American Journal of Intelligent Systems 2014, 4(2): 43-72
DOI: 10.5923/j.ajis.20140402.03

Multivariable NNARMAX Model Identification of an AS-WWTP Using ARLS: Part 1 – Dynamic Modeling of the Biological Reactors

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 five biological reactor units 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. The algorithms are validated by one-step and five-step ahead prediction methods. The performance of the two algorithms is assessed by using the Akaike's method to estimate the final prediction error (AFPE) of the regularized criterion. Furthermore, 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

Wastewater flows are highly nonlinear and its measurements in terms of volume are accurately performed graphically using some obtained data. Wastewater variations can then be obtained by processing these data statistically. In situations where adequate measurements of the volume of wastewater are not available, it has been proposed in [1] that estimates and calculation should be performed; and for this purpose the wastewater is divided into typical three (3) parts: domestic wastewater, industrial and public institutions wastewater, and infiltration. Different analytical methods of a mixed origin are used for the characterization of wastewater and sludge as presented in [1], and many of them have been specially developed for treatment plants and treatment processes. Wastewater normally contains thousands of different organics and a measurement of each individual organic matter would be impossible rather a different collective analyses are used which comprise a greater or minor part of the organics. Activated sludge wastewater treatment processes (ASWWTP) are difficult to control because of their complexity; nonlinear behaviour; large uncertainty in uncontrolled inputs and in the model parameters and structure; multiple time scales of the dynamics and multivariable input-output structure. The first control opportunity in ASWWTP is regulating the influent flow-rate which implies that control issues in wastewater treatment facilities pertain primarily to aeration control for energy usage and satisfying process demands.

The most straightforward controller design techniques are those that are based on a linear mathematical model of the controlled process [2]. However, the characteristics of the ASWWTP processes are highly nonlinear and time-varying in nature. In these cases the linear controller design techniques result to inefficient control algorithms [3] and methods based on nonlinear models of the processes are preferred [4], [5]. In either of the linear or nonlinear cases, the use of a model of the process does not fully reflect the actual process operation over long periods of time. Therefore, the algorithms obtained by controller design techniques which are based on a mathematical model of the controlled process [2] are not very efficient because these methods cannot guarantee stable control outside the range of the model validity [3], [6]. For these reasons adaptive algorithms which could be based on a continuous model updating process and redesign of the control strategy before a new control action is applied to the real plant would result to a better plant performance.

Even if this approach is adopted, the application of

* Corresponding author:
avincent@ieee.org (Vincent A. Akpan)
Published online at http://journal.sapub.org/ajis

traditional modeling methods used in several variations of the controller designs, such as those reported in [7]–[9], cannot model accurately the strong interactions among the process variables as well as the short and tight operating constraints. The best approach would be the use of highly complicated validated models of groups of nonlinear differential and partial differential equations, and the invention of new control design methods based on these models. However the computational burden for modeling dynamic systems with relatively short sampling interval becomes enormous to be handled even by the new multi-core, clustering and field programmable gate array (FPGA) technologies. In order to exploit these technologies, instead of using groups of differential equations, one could consider developing other accurate nonlinear models, the computational burden of which would be of course higher than the linear models but less than that of the groups of differential equations.

A recent approach to modeling nonlinear dynamical systems is the use of neural networks (NN). The application of neural networks (NN) for model identification and adaptive control of dynamic systems has been studied extensively [6], [10]–[19]. As demonstrated in [13], [14], [16], and [17], neural networks can approximate any nonlinear function to an arbitrary high degree of accuracy. The adjustment of the NN parameters results in different shaped nonlinearities achieved through a gradient descent approach on an error function that measures the difference between the output of the NN and the output of the true system for given input data or input-output data pairs (training data).

In the absence of operating data from the transient and steady state operation of the system to be controlled, data for training and testing the NN model can be obtained from the system by simulating the validated model of the groups of differential equations which are usually derived from the first principles on which the operation of the physical process is based. Such approaches are reported in [6], [10], 18], [21]. The use of the nonlinear NN models can replace the first principles model equally well and it can reduce the computational burden as argued in [21], [22]. This is because a nonlinear discrete NN model of high accuracy is available immediately after or at each instant of the network training process.

The aim of the paper is on the efficient modeling of a nonlinear neural network autoregressive moving average with exogenous input (NNARMAX) model that will capture the nonlinear dynamics of an activated sludge wastewater treatment plant (AS-WWTP) 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 present paper (Part 1) considers the modeling of the five biological reactors including the influent tank while the second paper (Part 2) will be devoted to the modeling of the secondary settler and clarifier as well as the effluent tank.

2. The AS-WWTP Process Description

2.1. Overview of the AS-WWTP Process

The activated sludge process was developed in England in 1914 by Arden and Lockett (Arden, et al.) and was so named because in involved the production of an activated mass of microorganisms capable of aerobically stabilizing a waste. The activated sludge process has been utilized for treatment of both domestic and industrial wastewaters for over half a century. This process originated from the observation made a long time ago that whenever wastewater, either domestic or industrial, is aerated for a period of time, the content of organic matter is reduced, and at the same time a flocculent sludge is formed. Microscopic examination of this sludge reveals that it is formed by a heterogenous population of microorganisms, which changes continually in nature in response to variation in the composition of the wastewater and environmental conditions. Microorganisms present are unicellular bacteria, fungi, algae, protozoa, and rotifiers.

Organic wastewater is introduced into a reactor where an aerobic bacterial culture is maintained in suspension. The reactor contents are referred to as the mixed liquor. In the reactor, the bacterial culture converts the organic content of the wastewater into cell tissue. The aerobic environment in the reactor is achieved by the use of diffused or mechanical aeration, which also serve to maintain the liquor in a completely mixed regime. After a specified period of time, the mixture of new cells and old cells is passed into a settling tank where the cells are separated from the treated wastewater. A portion of the settled cells is recycled to maintain the desired concentration of organisms in the reactor, and a portion is wasted.

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 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, ASTP can achieve biological nitrogen removal, biological phosphorus removal and removal of organic carbon substances as well as the amount of dissolved oxygen [23]. Generally, an ASWWTP can generally be regarded as a complex system [24]–[26] due to its highly nonlinear dynamics, large uncertainty, multiple time scales in the internal processes as well as its multivariable structure. A widely accepted biodegradation model is the activated sludge model no. 1 (ASM1) which incorporates the basic biotransformation processes of an activated sludge wastewater treatment plant [1], [27]–[30].

The objective of the activated sludge process is to achieve a sufficiently low concentration of biodegradable matter in the effluent together with minimal sludge production at

minimal cost. Although the operation of the ASP for wastewater treatment plants is challenging for both economical and technical reasons but the basic principle in activated sludge plants is that a mass of activated sludge is kept moving in water by stirring or aeration. Apart from the living biomass, the suspended solids contain inorganic as well as organic particles. While some of the organic particles can be degraded through hydrolysis, the others are non-degradable (inert). The amount of suspended solids in the treatment plant is regulated through recycle of the suspended solids and by removing the so-called excess sludge. The purpose of recycle is to increase the sludge concentration in the aeration tank which results in an increase sludge age as the hydraulic retention time and the sludge age are separated. In this way it is possible to accumulate a biomass consisting of both rapid- and slow-growing micro-organisms which is very important for the settling and flocculation behaviour of the sludge.

The activated sludge process is a treatment technique in which wastewater and reused biological sludge full of living microorganisms are mixed and aerated. The biological solids are then separated from the treated wastewater in a clarifier and are returned to the aeration process or wasted. The microorganisms are mixed thoroughly with the incoming organic material, and they grow and reproduce by using the organic material as food. As they grow and are mixed with air, the individual organisms clings together (flocculate). Once flocculated, they more readily settle in the secondary clarifiers. The wastewater being treated flows continuously into an aeration tank where air is injected to mix the wastewater with the returned activated sludge and to supply the oxygen needed by the microbes to live and feed on the organics. Aeration can be supplied by injection through air diffusers in the bottom of tank or by mechanical aerators located at the surface.

The mixture of activated sludge and wastewater in the aeration tank, mixed liquor, flows to a secondary 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 an activated sludge system. 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 sludge age. It should be noted 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.

2.2. Description of the AS-WWTP Process

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. 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] 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 Appendix A. 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 complete BSM1 used to describe the AS-WWTP considered here is given in Appendix A while the general characteristics of the biological reactors are given in Appendix B. Additional information on the complete mathematical modeling of the AS-WWTP considered here can be found in [27].

The schematic of a BNR-ASWWTP design with basic control strategies is shown in Fig.1 using the *Johannesburg* configuration [24], [25] which consists of anaerobic, anoxic and aerobic zones and a secondary settler in a back-to-back scheme with multiple recycle streams [27]. 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 is detailed in Appendix C of [27]. Nevertheless, the biological processes within the different zones of the reactors are briefly presented in the following.

In the anaerobic zone, fermentable organic from the influent wastewater are mixed with the return-activated sludge (RAS) and converted to volatile fatty acids (VFA) by heterotrophic organisms. The latter is consumed by phosphorus-accumulating organisms (PAO) and stored internally as poly-β hydroxy alkanoates (PHA). Concurrently, poly phosphate and hence energy for VFA accumulation are internally released. Denitrification in this zone results in a net reduction of alkalinity and hence there is an increase in pH due to acids production. If the amount of VFA is insufficient, additional acids from external source may be added to maintain a maximum PHA uptake by the biological phosphate organisms. It is also common to install an activated primary sedimentation tank to allow production of VFA by fermentation of readily substrate in the incoming sewage.

In the anoxic zone, nitrate (S_{NO}) which is recycled from the aerobic zone is converted to dinitrogen by facultative heterotrophic organisms. Denitrification in this zone results in the release of alkalinity and hence there is an increase in pH value. There is also evidence of a pronounced removal of phosphorus in this zone.

In the partially-treated wastewater arriving the aerobic zone, virtually al the readily biodegradable organic (referred to as biodegradable COD) in the partially-treated wastewater has been consumed by heterotrophic organisms in the aerobic and anoxic zones. Thus in this aerobic zone, two major processes occur. In the presence of dissolved oxygen (S_O), the released phosphate is taken up by PAO growing on the stored PHA. The phosphorus is stored internally as poly phosphate. This results in a net reduction in phosphate in the wastewater. The second process occurring in this zone is nitrification of ammonia to nitrate in the wastewater by the autotrophic organisms. In order to minimize the amount of DO going into the anoxic zone, the last compartment is typically aerated. Part of the sludge, which contains phosphorus to be removed, is wasted while the remainder is returned to the anaerobic zone after thickening in the settler and additional denitrification in the RAS tank.

The activated sludge wastewater treatment plant considered here is strictly based on the benchmark simulation model no. 1 (BSM1) 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) [28], [29]. This implementation of the benchmark simulation model no. 1 (BSM1) follows the methodology specified in [27]–[29] 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) is given in Appendix C of [27] 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 BSM1.

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

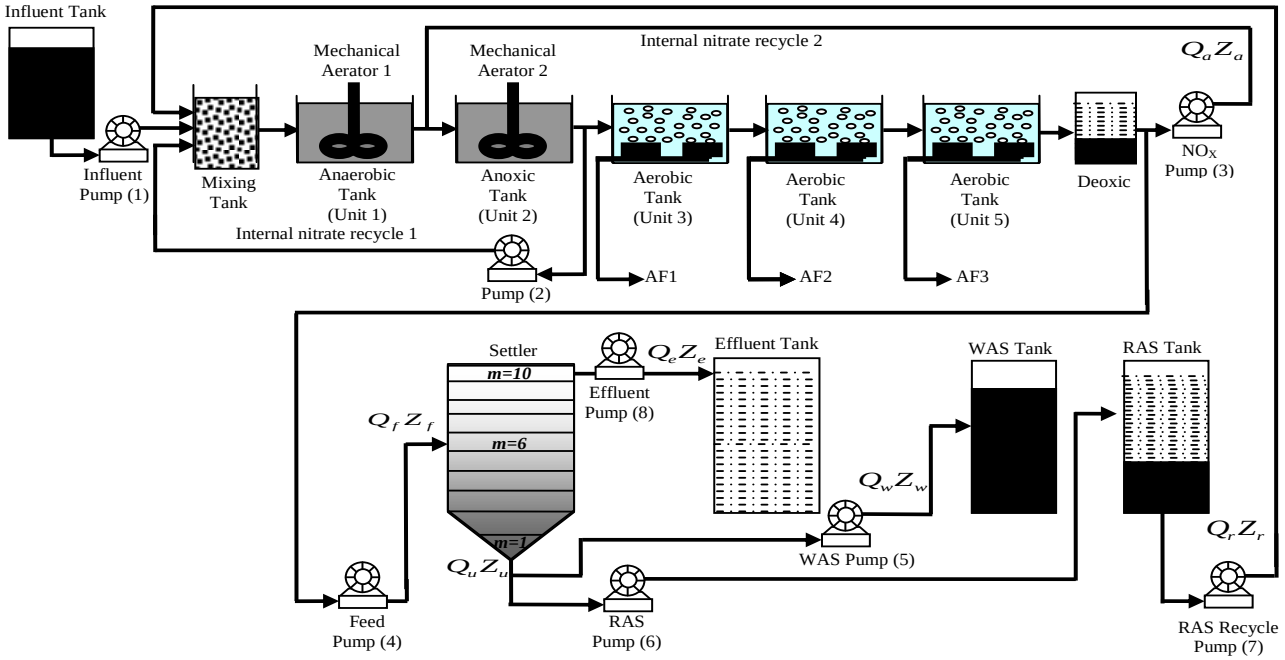


Figure 1. The schematic of the AS-WWTP process

3. The Neural Network Identification Scheme and Validation Algorithms

3.1. 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 [30], [31] and control [12], [13], [32], [33] 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 [30], [31]:

$$\left.\begin{aligned} Y(k) = J\big[&Y(k-1),\ldots,Y(k-n),\\ &U(k-d),\ldots,U(k-d-m)\big]+\tilde{d}(k)\end{aligned}\right\} \quad (1)$$

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 [30], [31]:

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

where

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

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

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

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

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

where $\mathbb{Q}_\theta$ is some subset of $\Re^\nu$ 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:

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

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

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

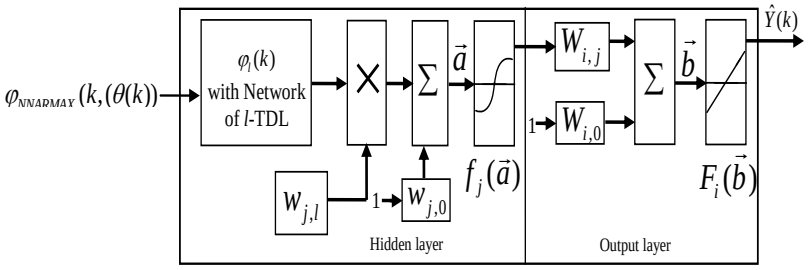
where $J(Z^N, \varphi(k,\theta(k)), \theta(k))$ is formulated as a total square error (TSE) type cost function which can be stated as:

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

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 $J(Z^N, \varphi(k,\theta(k)), \theta(k))$.

3.2. Neural Network Identification Scheme

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 $\varphi_{l_m}(k) = [U(k-d),\ldots,U(k-d-m)]$, $\varphi_{l_n}(k) = [Y(k-1),\ldots,Y(k-n)]^T$ and $\varphi_{l_e}(k,\theta(k)) = [\varepsilon(k-1,\theta(k)),\ldots,\varepsilon(k-n,\theta(k))]^T$ 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:



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

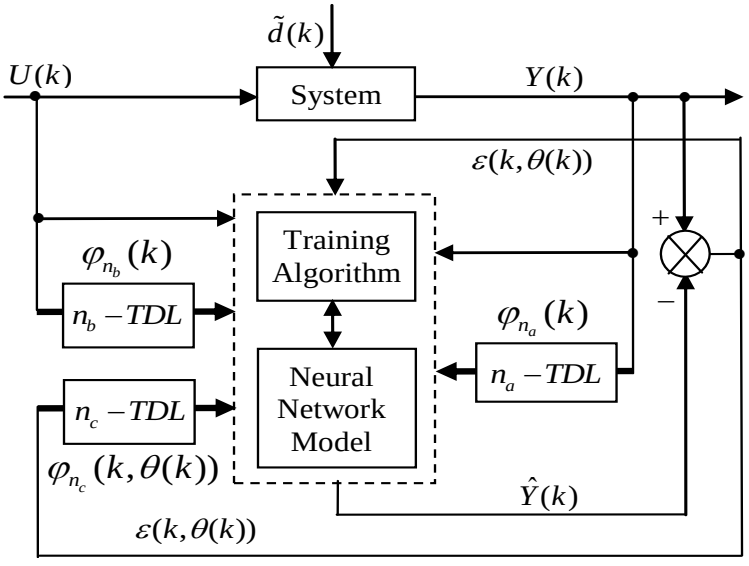


Figure 3. NN model identification based on the teacher-forcing method

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

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 $\mathbf{w} = \{w_{j,l} \;\; w_{j,o}\}$ 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:

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

where the second term in (10) is the regularization (weight decay) term [31] 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)*.

3.3. 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 [30]–[34]:

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

where

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

$\theta_\tau(k)$ 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

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

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

$$\left.\begin{array}{l}\psi[k, \theta_\tau(k)] = \dfrac{\partial\hat{Y}(k \mid \theta(k))}{\partial\theta(k)} \\ \qquad - \dfrac{\partial\hat{Y}(k \mid \theta(k))}{\partial\varepsilon(k-1, \theta(k))} \dfrac{\partial\hat{Y}(k-1 \mid \theta(k))}{\partial\theta(k)} \\ \qquad - \cdots - \dfrac{\partial\hat{Y}(k \mid \theta(k))}{\partial\varepsilon(k-n_c, \theta(k))} \dfrac{\partial\hat{Y}(k-n_c \mid \theta(k))}{\partial\theta(k)}\end{array}\right\} \tag{14}$$

Thus, Equation (14) can be expressed equivalently as

$$\left.\begin{array}{l}\psi[l, \theta(k)] = \dfrac{d\hat{Y}(k \mid \theta(k))}{d\theta(k)} - C_1(k)\psi[k-1, \theta(k)] - \\ \qquad \cdots - C_n(k)\psi[k-n, \theta(k)]\end{array}\right\} \tag{15}$$

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

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

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.

3.3.1. 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 [27], [35]:

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

$\pi \in (0,1)$ 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:

$$\hat{\theta}(k) = \theta_\tau(k\,|\,k-1) - R\left[\theta_\tau(k\,|\,k-1)\right]^{-1} G\left[\theta_\tau(k\,|\,k-1)\right] \quad (18)$$

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

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

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

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

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

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

$$\left.\begin{aligned} P(k) &= \frac{1}{\pi}P(k-1) - \Lambda(k)\psi^T[l,\theta_\tau(k-1)]P(k-1)\\ &+\beta I - \delta' P^2(k-1)\end{aligned}\right\} \quad (21)$$

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 [27], [35], [36]:

$$\left.\begin{aligned} &0<\gamma<\alpha<1, \qquad \beta>0, \qquad \delta'>0,\\ &(\gamma-\alpha)^2 + 4\beta\delta' < (1-\alpha)^2\end{aligned}\right\} \quad (22)$$

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 $\bar{e}_{\min}$ and maximum $\bar{e}_{\max}$ eigenvalues of (21) given respectively as [27], [35], [36]:

$$\left.\begin{aligned} \bar{e}_{\min} &= \left((\alpha-\gamma)/2\delta'\right)\left(-1+\sqrt{1+\left(4\beta\delta'/(\alpha-\gamma)^2\right)}\right)\\ \bar{e}_{\max} &= \left(\gamma/2\delta'\right)\left(1+\sqrt{1+\left(4\beta\delta'/\gamma^2\right)}\right)\end{aligned}\right\} \quad (23)$$

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

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

$$\left.\begin{aligned} \hat{\theta}(k) &= \theta_\tau(k-1)\\ &+\Lambda(k)\Gamma^{-1}[\theta(k)][Y(k) - \hat{Y}(k\,|\,\theta_\tau(k-1))]\end{aligned}\right\} \quad (24)$$

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.

3.4. Proposed Validation Methods for the Trained NNARMAX Model

Network validations are performed to assess to what extend the trained network captures and represents the operation of the underlying system dynamics [31], [34].

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

The second validation test is the Akaike's final prediction error (AFPE) estimate [31], [34] 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 [31], [36]:

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

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:

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

where

$$\hat{\varphi}(k+K) = [U((k+K-1)\,|\,\hat{\theta}),\ldots,\ U((k+K-m)\,|\,\hat{\theta}),$$
$$\hat{Y}((k+K-1)\,|\,\hat{\theta}),\ldots,\hat{Y}((k+K+1-\min(k,n))\,|\,\hat{\theta}),$$
$$Y((k+K-1)\,|\,\hat{\theta}),\ldots,Y((k+K-\max(n-k,0)\,|\,\hat{\theta})]^T$$

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:

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

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

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

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

Case I: Nonlinear Model Identification of the Anaerobic and Anoxic Reactors

This section concentrates on the nitrification and denitrification processes with focus on the activated sludge, food-to-microorganisms ratio as well as the recycled nitrates and sludge. The proposed identification and control strategy based on ANN is illustrated by the first two reactors shown in the upper segment of Fig.4.

1). Anaerobic Reactor (Unit 1):

The nitrification of ammonia into nitrites and nitrates occurs in this reactor and thus the objective here is to control the control the nitrate concentration (SNO1 = 3.5 $g.m^{-3}$) by manipulating the influent flow rate (Q_{IN} = 18446 $m^3.d^{-1}$), RAS recycle flow rate (Q_{R1} = 18446 $m^3.d^{-1}$), and the aeration intensity (K_La1 = 1 hr^{-1}) by using Salk1, SNH1, XND1, and XP1 as inputs for this control action.

2). Anoxic Reactor (Unit 2)

The denitrification of nitrates into atmospheric nitrogen by microorganisms takes place in this unit. The objectives here are to maintain SND2 = 1 g/m^3, XP2 = 448 $g.m^{-1}$, SNO2 = 3.6 $g.m^{-1}$ by manipulating the internal recycled nitrate 1 flow rate (Q_{A1} = 16485 $m^3.d^{-1}$) and the aeration intensity K_La2 = 2 hr^{-1} by using SNO2_measured and Salk2 as inputs.

Case II: Nonlinear Model Identification of the Aerobic Reactors

This section concentrates on the removal of biological nutrients, nitrogen and phosphorus to improve the quality of the expected effluent by manipulating the aeration intensities of the air flow (AF1, AF2 and AF3) and flow rates (Q_{A2} and Q_{F1}) by regulating the dissolved oxygen concentrations based on the biological nutrient, phosphorus and nitrogen concentrations. The proposed identification and control strategy based on ANN is illustrated by the last three aerobic reactors shown in the lower segment of Fig. 4.

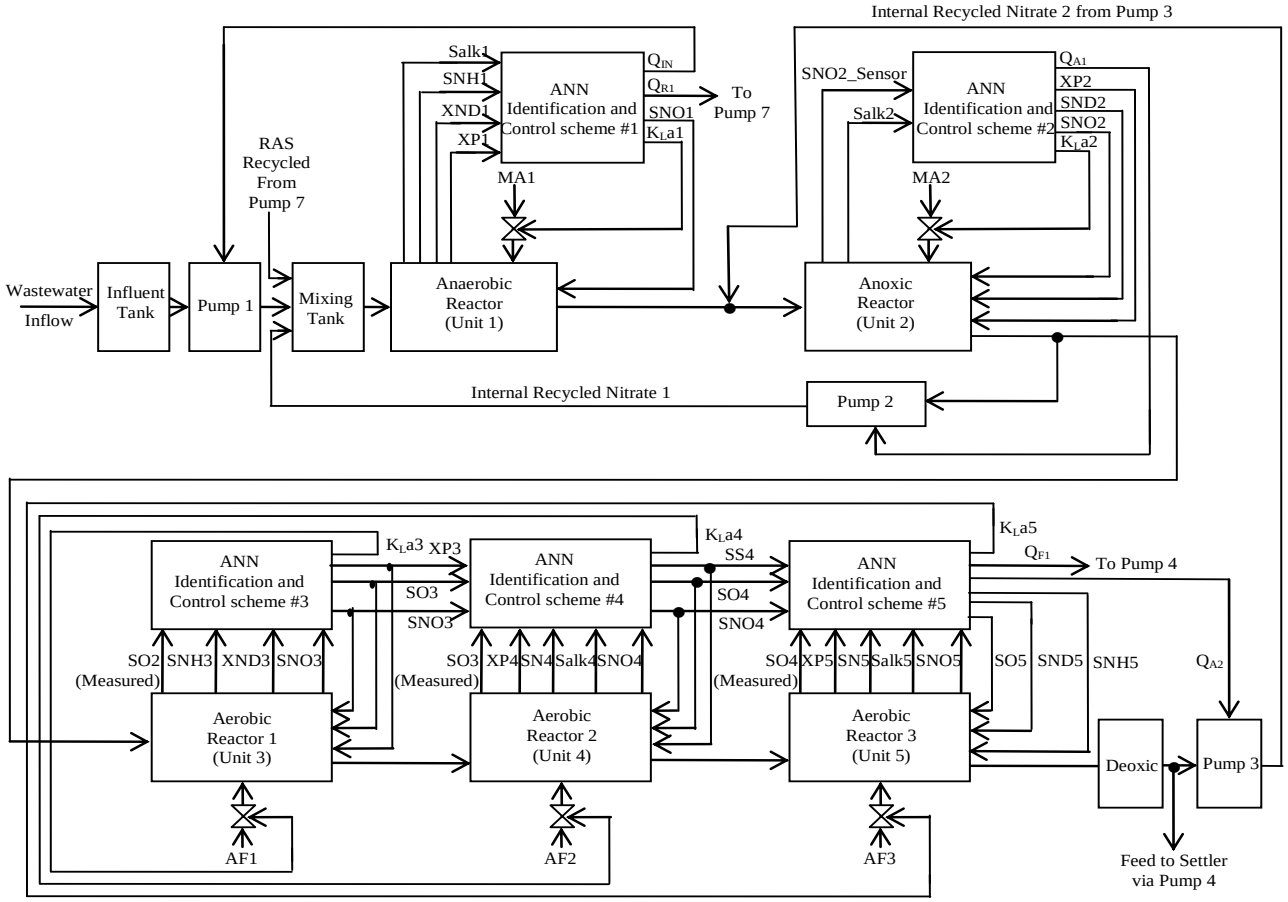


Figure 4. The multivariable neural network-based NNARMAX model identification scheme for the five biological reactors of the AS-WWTP with the proposed manipulated inputs and controlled outputs

3). First Aerobic Reactor (Unit 3)

The objective here is to regulate XP3 = 449 $g.m^{-3}$, SO3 = 2 $g.m^{-3}$ and SNO3 = 6.2 $g.m^{-1}$ by manipulating the aeration intensity (K_La3 = 10 hr^{-1} using SO2 (measured), SNH3, XND3, and SNO3 as inputs.

4). Second Aerobic Reactor (Unit 4)

The objective here is to regulate SO4 = 2 $g.m^{-3}$, SNO4 = 11.6 $g.m^{-1}$ and SS4 = 35 $g.m^{-3}$ by manipulating K_La4 = 10 hr^{-1} using SO3 (measured), XP4, SN4, Salk4, and SNO4 as inputs.

5). Third Aerobic Reactor (Unit 5)

The objective here is to regulate SNH5 = 35 $g.m^{-3}$, SND5 = 13.5 $g.m^{-1}$, SO5 = 2 $g.m^{-3}$ by manipulating the internal recycled nitrate 2 (Q_{A2} = 16485 $m^3.d^{-1}$), feed flow rate (Q_{F1} = 36892 $m^3.d^{-1}$) depending on the feed flow decision system, and the aeration intensity (K_La5 = 3.5 hr^{-1}) using SO4 (measured), XP5, SN5, Salk5, and SNO5 as inputs.

4.2. Formulation of the AS-WWTP Model Identification Problem

4.2.1. 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 as follows: S_I, S_S, X_I, X_S, X_{BH}, X_{BA}, X_P, S_O, S_{NO}, S_{NH}, S_{ND}, X_{ND}, and S_{ALK} out of which four states are measurable namely: S_S (readily biodegradable substrate), X_{BH} (active heterotrophic biomass), S_O (oxygen) and S_{NO} (nitrate and nitrite nitrogen). An additional important parameter TSS is used to assess the amount of soluble solids in all the reactors including aerobic reactor of Unit 5.

As highlighted above, the main objective here is on the efficient neural network model identification to obtain a multivariable NNARMAX model equivalent of the activated sludge wastewater treatment plant (AS-WWTP) with a view in using the obtained model for multivariable adaptive predictive control of the AS-WWTP process in our future work. Thus, from Section 2, the measured inputs that influence the behaviour of the AS-WWTP process shown in Fig. 5 are:

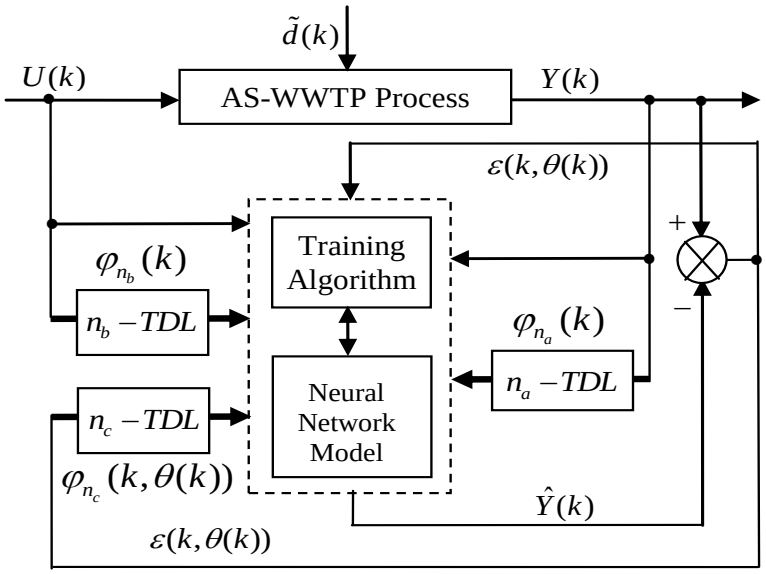


Figure 5. The neural network model identification scheme for AS-WWTP based on NNARMAX model

$$\left.\begin{aligned} U(k) = \big[& S_{alk1}(k) \;\; S_{NH1}(k) \;\; X_{ND1}(k) \;\; X_{P1}(k); \;\; S_{NO2}(k) \\ & S_{alk2}(k); \;\; S_{O2}(k) \;\; S_{NH3}(k) \;\; X_{ND3}(k) \;\; S_{NO3}(k); \\ & S_{O3}(k) \;\; X_{P4}(k) \;\; S_{N4}(k) \;\; S_{alk4}(k) \;\; S_{NO4}(k); \\ & S_{O4}(k) \;\; X_{P5}(k) \;\; S_{N5}(k) \;\; S_{alk5}(k) \;\; S_{NO5}(k) \big]^T \end{aligned}\right\} \tag{28}$$

From the arguments in Section 2, the output parameters that capture the behaviour of the AS-WWTP are defined here as:

$$\left.\begin{aligned} Y(k) = \big[& y_{SI}(k) \;\; y_{SS}(k) \;\; y_{XI}(k); \;\; y_{XS}(k) \;\; y_{XBH}(k) \\ & y_{XBA}(k) \;\; y_{XP}(k) \;\; y_{SO}(k) \;\; y_{SNO}(k) \;\; y_{SNH}(k) \\ & y_{SND}(k) \;\; y_{XND}(k) \;\; y_{SALK}(k) \;\; y_{TSS}(k); \;\; y_{KLa1}(k) \\ & y_{Q\inf}(k) \;\; y_{Qras}(k) \;\; y_{SNO1}(k); \;\; y_{SNO2}(k) \;\; y_{XP2}(k) \\ & y_{Qirf}(k) \;\; y_{KLa2}(k) \;\; y_{SND2}(k); y_{SO3}(k) \;\; y_{KLa3}(k) \\ & y_{SNO3}(k) \;\; y_{XP3}(k); \;\; y_{SO4}(k) \;\; y_{KLa4}(k) \;\; y_{SNO4}(k) \\ & y_{SS4}(k); \;\; y_{SO5}(k) \;\; y_{KLa5}(k) \;\; y_{Qirn}(k) \;\; y_{Qffr}(k) \\ & y_{N5}(k) \;\; y_{SNH5}(k) \big]^T \end{aligned}\right\} \tag{29}$$

Although, the multivariable system is formulated as 26–input 23–output control problem, but the neural network model identification is a much more complicated multiple–input multiple–output (MIMO) problem since all the fourteen states must be predicted at each sampling instant in order to obtain a reasonable approximate model that describes and captures the system's dynamics at that instant. Thus, making the total outputs 37. 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 3. 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 follows:

$$\begin{aligned}\varphi_{n_a}(k) = \big[& S_I(k-n_a) \;\; S_S(k-n_a) \;\; X_I(k-n_a) \\ & X_S(k-n_a) \;\; X_{BH}(k-n_a) \;\; X_{BA}(k-n_a) \\ & X_P(k-n_a) \;\; S_O(k-n_a) \;\; S_{NO}(k-n_a) \\ & S_{NH}(k-n_a) \;\; S_{ND}(k-n_a) \;\; X_{ND}(k-n_a) \\ & S_{ALK}(k-n_a) \;\; TSS(k-n_a); \;\; K_{La1}(k-n_a) \\ & Q_{\inf}(k-n_a) \;\; Q_{ras}(k-n_a) \;\; S_{NO1}(k-n_a); \\ & S_{NO2}(k-n_a) \;\; X_{P2}(k-n_a) \;\; Q_{irf}(k-n_a) \\ & K_{La2}(k-n_a) \;\; S_{ND2}(k-n_a); \;\; S_{O3}(k-n_a) \\ & K_{La3}(k-n_a) \;\; S_{NO3}(k-n_a) \;\; X_{P3}(k-n_a); \\ & S_{O4}(k-n_a) \;\; K_{La4}(k-n_a) \;\; S_{NO4}(k-n_a) \\ & S_{S4}(k-n_a); \;\; S_{O5}(k-n_a) \;\; K_{La5}(k-n_a) \\ & Q_{irn}(k-n_a) \;\; Q_{ff}r(k-n_a) \;\; S_{N5}(k-n_a) \\ & S_{NH5}(k-n_a)\big]^T \end{aligned} \tag{30}$$

$$\begin{aligned}\varphi_{n_b}(k) = \big[& S_{alk1}(k-n_b) \;\; S_{NH1}(k-n_b) \;\; X_{ND1}(k-n_b) \\ & X_{P1}(k-n_b); \;\; S_{NO2}(k-n_b) \;\; S_{alk2}(k-n_b); \\ & S_{O2}(k-n_b) \;\; S_{NH3}(k-n_b) \;\; X_{ND3}(k-n_b) \\ & S_{NO3}(k-n_b); \;\; S_{O3}(k-n_b) \;\; X_{P4}(k-n_b) \\ & S_{N4}(k-n_b) \;\; S_{alk4}(k-n_b) \;\; S_{NO4}(k-n_b); \\ & S_{O4}(k-n_b) \;\; X_{P5}(k-n_b) \;\; S_{N5}(k-n_b) \\ & S_{alk5}(k-n_b) \;\; S_{NO5}(k-n_b)\big]^T \end{aligned} \tag{31}$$

$$\begin{aligned}\varphi_{n_c}(k,\theta(k)) = \big[& \ldots \\ & \varepsilon_{S_I}(k-n_c,\theta(k)) \;\; \varepsilon_{S_S}(k-n_c,\theta(k)) \;\; \varepsilon_{X_I}(k-n_c,\theta(k)) \\ & \varepsilon_{X_S}(k-n_c,\theta(k)) \;\; \varepsilon_{X_{BH}}(k-n_c,\theta(k)) \;\; \varepsilon_{X_{BA}}(k-n_c,\theta(k)) \\ & \varepsilon_{X_P}(k-n_c,\theta(k)) \;\; \varepsilon_{S_O}(k-n_c,\theta(k)) \;\; \varepsilon_{S_{NO}}(k-n_c,\theta(k)) \\ & \varepsilon_{S_{NH}}(k-n_c,\theta(k)) \;\; \varepsilon_{S_{ND}}(k-n_c,\theta(k)) \;\; \varepsilon_{X_{ND}}(k-n_c,\theta(k)) \\ & \varepsilon_{S_{ALK}}(k-n_c,\theta(k)) \;\; \varepsilon_{TSS}(k-n_c,\theta(k)); \;\; \varepsilon_{KLa1}(k-n_c,\theta(k)) \\ & \varepsilon_{Q\inf}(k-n_c,\theta(k)) \;\; \varepsilon_{Qras}(k-n_c,\theta(k)) \;\; \varepsilon_{SNO1}(k-n_c,\theta(k)); \\ & \varepsilon_{SNO2}(k-n_c,\theta(k)) \;\; \varepsilon_{XP2}(k-n_c,\theta(k)) \;\; \varepsilon_{Qirf}(k-n_c,\theta(k)) \\ & \varepsilon_{KLa2}(k-n_c,\theta(k)) \;\; \varepsilon_{SND2}(k-n_c,\theta(k)); \;\; \varepsilon_{SO3}(k-n_c,\theta(k)) \\ & \varepsilon_{KLa3}(k-n_c,\theta(k)) \;\; \varepsilon_{SNO3}(k-n_c,\theta(k)) \;\; \varepsilon_{XP3}(k-n_c,\theta(k)); \\ & \varepsilon_{SO4}(k-n_c,\theta(k)) \;\; \varepsilon_{KLa4}(k-n_c,\theta(k)) \;\; \varepsilon_{SNO4}(k-n_c,\theta(k)) \\ & \varepsilon_{SS4}(k-n_c,\theta(k)); \;\; \varepsilon_{SO5}(k-n_c,\theta(k)) \;\; \varepsilon_{KLa5}(k-n_c,\theta(k)) \\ & \varepsilon_{Qirn}(k-n_c,\theta(k)) \;\; \varepsilon_{Qffr}(k-n_c,\theta(k)) \;\; \varepsilon_{N5}(k-n_c,\theta(k)) \\ & \varepsilon_{SNH5}(k-n_c,\theta(k))\big]^T \end{aligned} \tag{32}$$

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

The outputs of the neural network for the AS-WWTP process are the predicted values of the thirteen states together with the amount of total soluble solids (*TSS*), thus resulting in fourteen states to be predicted at each sampling instant given by:

$$\left.\begin{aligned} \hat{Y}(k) = \big[& \hat{y}_{SI}(k) \quad \hat{y}_{SS}(k) \quad \hat{y}_{XI}(k); \quad \hat{y}_{XS}(k) \quad \hat{y}_{XBH}(k) \\ & \hat{y}_{XBA}(k) \quad \hat{y}_{XP}(k) \quad \hat{y}_{SO}(k) \quad \hat{y}_{SNO}(k) \quad \hat{y}_{SNH}(k) \\ & \hat{y}_{SND}(k) \quad \hat{y}_{XND}(k) \quad \hat{y}_{SALK}(k) \quad \hat{y}_{TSS}(k); \quad \hat{y}_{KLa1}(k) \\ & \hat{y}_{Q\inf}(k) \quad \hat{y}_{Qras}(k) \quad \hat{y}_{SNO1}(k); \quad \hat{y}_{SNO2}(k) \quad \hat{y}_{XP2}(k) \\ & \hat{y}_{Qirf}(k) \quad \hat{y}_{KLa2}(k) \quad \hat{y}_{SND2}(k); \quad \hat{y}_{SO3}(k) \quad \hat{y}_{KLa3}(k) \\ & \hat{y}_{SNO3}(k) \quad \hat{y}_{XP3}(k); \quad \hat{y}_{SO4}(k) \quad \hat{y}_{KLa4}(k) \quad \hat{y}_{SNO4}(k) \\ & \hat{y}_{SS4}(k); \quad \hat{y}_{SO5}(k) \quad \hat{y}_{KLa5}(k) \quad \hat{y}_{Qirn}(k) \quad \hat{y}_{Qffr}(k) \\ & \hat{y}_{N5}(k) \quad \hat{y}_{SNH5}(k) \big]^T \end{aligned}\right\} \tag{34}$$

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 [27]–[29]. 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 in the last sub-section of this section to further investigate the closed-loop controllers' performances based on an updated neural network model at each sampling time instants.

4.2.2. 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.

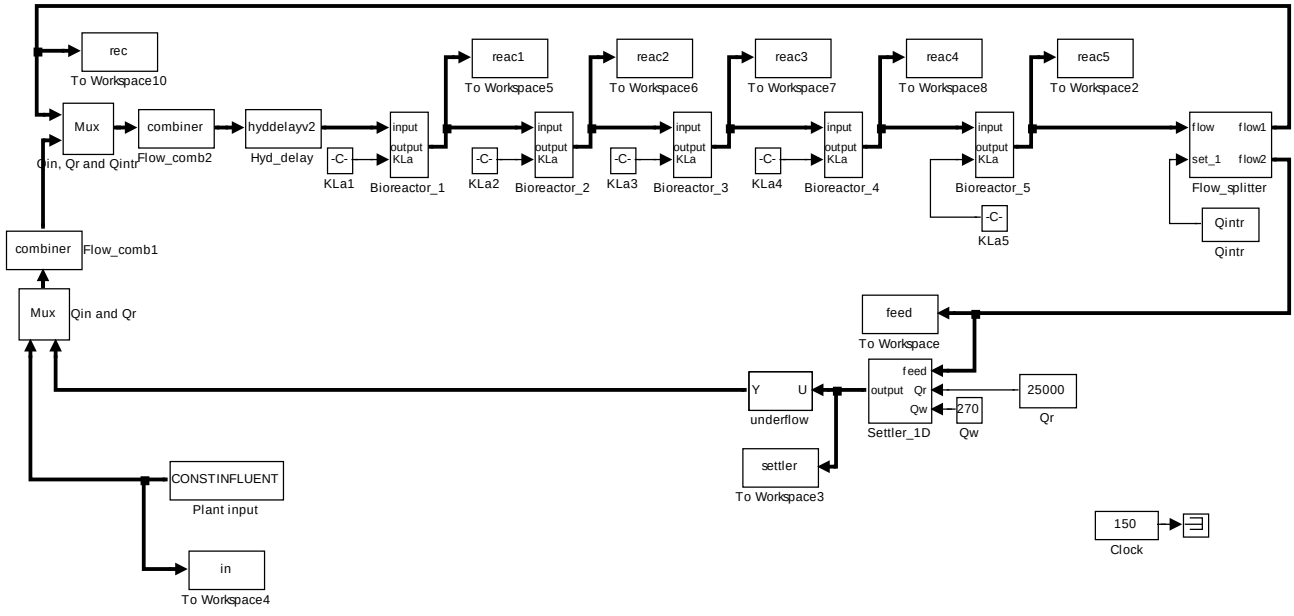


Figure 6. Open-loop steady-state benchmark simulation model No.1 (BSM1) with constant influent

A two-step simulation procedure defined in the simulation benchmark [27]–[29] 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 130first day dry-weather data samples provided by the COST Actions 624 and 682 is used for the trained NN validation.

4.2.3. 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 to this purpose. The incremental or online back-propagation (INCBP) algorithm was originally proposed by [38] which has been modified in [27] 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 3.3.1under the formulation of the ARLS algorithm; that is:

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

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).

4.2.4. 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:

$$\left.\begin{aligned}U^{(s)}(k)&=\frac{U(k)-\bar{U}(k)}{\sigma_{U(k)}}\\ Y^{(S)}(k)&=\frac{Y(k)-\bar{Y}(k)}{\sigma_{Y(k)}}\end{aligned}\right\} \tag{37}$$

where $\bar{U}(k)$, $\bar{Y}(k)$ 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

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

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.

4.2.5. Training the Neural Network that Models the Biological Reactors of the AS-WWTP Process

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 τ = 100 epochs (number of iterations) with the following selected parameters: $p=26$, $q=37$, $n_a=2$, $n_b=2$, $n_c=2$, $n_\varphi=300$ (NNARMAX), $n_h=5$, $n_o=37$, $\alpha_h=1e-6$ 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).

Table 4. Anaerobic reactor (Unit 1)

	K_La_1		QIN (Qinf)		Q_{ras}		SNO_1	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	1.4102e+001	1.2599e+002	1.4352e+001	1.2588e+002	1.4617e+001	5.7096e+001	1.3666e+001	1.2728e+002
Total square error (TSE)	6.7750e-001	3.7256e-001	4.1244e+003	1.4150e+003	5.2800e-001	4.2855e-003	4.9809e-001	8.2568e-003
Minimum performance index	2.9139e-002	6.1142e-003	1.0592e-001	2.5017e-003	3.7097e-002	6.2664e-006	5.2736e-002	7.5376e-006
Mean error of one step ahead prediction of training data	4.3949e-002	3.5303e-003	6.0938e+001	2.8049e+001	1.1611e-003	3.7946e-004	7.5569e-003	1.4868e-004
Mean error of one step prediction of test data	2.6306e-003	3.6821e-004	2.1687e-002	2.3878e-004	3.9983e-003	7.3588e-005	5.3524e-002	1.1808e-005
Mean value of 5-step ahead prediction error	4.4733e+000	3.5957e+000	5.4352e+000	4.7216e+000	5.4352e+000	4.7216e+000	5.4352e+000	4.7216e+000
Akaike's final prediction error (AFPE) estimate	2.9263e-002	6.1525e-003	7.7825e+006	1.8386e+005	2.1149e-002	3.5650e-006	3.1525e-002	4.4149e-006

Table 5(a). Anoxic reactor (Unit 2)

	SNO_2		XP_2		QA1 (Qirf)	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	1.5538e+001	1.3595e+002	1.5038e+001	1.4026e+002	1.3728e+001	1.2599e+002
Total square error (TSE)	2.0856e-001	9.9078e-003	1.2669e+001	1.4709e+000	6.8958e+003	3.8328e+003
Minimum performance index	1.0495e-002	2.7919e-005	1.7274e-002	1.2377e-004	2.6931e-002	4.4129e-003
Mean error of one step ahead prediction of training data	2.8533e-002	1.4300e-003	2.7607e+000	1.5838e-001	5.3938e+001	1.0736e+000
Mean error of one step prediction of test data	3.4418e-003	1.6147e-004	2.4235e-002	2.6206e-005	4.3678e-002	4.6612e-004
Mean value of 5-step ahead prediction error	1.8773e+001	1.8933e+001	9.6140e-001	1.5754e+000	Infinite	Infinite
Akaike's final prediction error (AFPE) estimate	2.6585e-003	7.0709e-006	9.5517e+001	6.8547e-001	8.7867e+006	1.4445e+006

Table 5(b). Anoxic reactor (Unit 2)

	K_La_2		SND_2	
	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	1.3837e+001	6.7564e+001	1.4336e+001	3.4437e+002
Total square error (TSE)	6.8729e-001	8.4789e-003	1.1638e-001	3.1700e-002
Minimum performance index	5.6393e-002	6.8074e-006	5.2053e-002	1.4765e-003
Mean error of one step ahead prediction of training data	5.8302e-002	2.3397e-004	5.1565e-003	1.4663e-004
Mean error of one step prediction of test data	1.5618e-002	1.6080e-005	3.8503e-002	6.3743e-005
Mean value of 5-step ahead prediction error	Infinite	Infinite	4.4573e-001	2.4749e+000
Akaike's final prediction error (AFPE) estimate	3.1998e-002	3.8212e-006	3.4321e-003	9.7738e-005

Table 6. First aerobic reactor (Unit 3)

	SO_3		K_La_3		SNO_3		XP_3	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	3.6972e+001	3.3790e+002	3.9780e+000	1.6209e+001	3.8220e+000	1.6224e+001	4.6020e+000	3.1247e+001
Total square error (TSE)	1.6848e-001	3.2068e-002	2.6009e-001	1.1328e-002	7.3386e-002	1.1517e-001	7.4634e+000	1.1247e+000
Minimum performance index	7.6041e-003	4.4054e-004	2.7946e-001	3.4834e-005	4.2569e-00	4.4553e-004	1.5452e-001	7.1342e-005
Mean error of one step ahead prediction of training data	2.8745e-003	3.5692e-003	3.2986e-001	3.0566e-004	1.3681e+000	5.7661e-003	3.2373e+001	1.9295e-002
Mean error of one step prediction of test data	3.5020e-003	6.2088e-005	2.9852e-001	4.4456e-005	2.5590e-001	6.3008e-005	4.3663e-001	7.1683e-005
Mean value of 5-step ahead prediction error	1.0667e+000	1.1345e+000	3.3534e+000	9.2814e-002	2.9939e+000	3.5237e+000	1.8289e+001	6.5787e-001
Akaike's final prediction error (AFPE) estimate	2.9988e-003	1.7452e-004	1.4468e-001	1.8217e-005	1.4070e+000	1.5022e-003	8.3644e+002	3.8383e-001

Table 7. Second aerobic reactor (Unit 4)

	SO_4		K_La_4		SNO_4		SS_4	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	3.9156e+000	1.4586e+001	3.6192e+000	1.7706e+001	3.6816e+000	1.5522e+001	3.4944e+000	3.1013e+001
Total square error (TSE)	5.0251e-001	3.5628e-002	4.0161e-001	1.2237e-002	4.0775e-001	2.5905e-001	7.2445e-001	7.3863e-003
Minimum performance index	3.6716e-002	1.1234e-004	7.5769e-002	2.8757e-005	4.3564e-001	1.5645e-004	6.2008e-001	2.3445e-004
Mean error of one step ahead prediction of training data	4.3352e-002	5.4949e-004	3.7356e-002	8.3317e-004	4.3352e-002	5.4949e-004	3.7356e-002	8.3317e-004
Mean error of one step prediction of test data	3.2220e-002	1.9972e-004	1.7794e-001	1.7722e-004	1.1457e-001	1.1457e-001	4.2808e-001	3.6860e-005
Mean value of 5-step ahead prediction error	1.1986e-001	1.1173e+000	2.9488e+000	1.0770e-001	1.0513e+001	3.0935e+000	2.9845e+001	3.8497e+000
Akaike's final prediction error (AFPE) estimate	2.6339e-002	7.8441e-005	4.1964e-002	1.5593e-005	3.5485e+000	1.2913e-003	1.0790e-001	3.9854e-005

Table 8(a). Third aerobic reactor (Unit 5)

	SO_5		K_La_5		QA2 (Q_{irn})		QF1 (Q_{ffr})	
	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	1.4960e+001	1.2931e+002	7.7064e+000	3.5319e+001	7.3164e+000	3.5225e+001	8.3149e+000	7.4428e+001
Total square error (TSE)	1.9376e-001	1.1617e-001	3.8041e-001	1.4410e-002	3.1299e+004	5.4457e+003	3.1299e+004	5.4457e+003
Minimum performance index	8.6572e-002	1.6042e-002	3.4415e-001	2.1632e-005	6.8819e-001	9.0292e-003	3.6145e-001	1.7517e-002
Mean error of one step ahead prediction of training data	7.9349e-004	2.8091e-004	3.5338e-001	3.7364e-004	2.1703e+004	7.0495e+001	3.6456e+003	1.9609e+002
Mean error of one step prediction of test data	9.1350e-003	4.0413e-003	4.6217e-001	1.6773e-004	3.7270e-001	3.2919e-005	7.3937e-002	9.9085e-004
Mean value of 5-step ahead prediction error	2.3612e-001	2.0044e-001	3.7306e+000	1.1056e-001	Inf	Inf	8.6831e+000	2.8156e+000
Akaike's final prediction error (AFPE) estimate	8.0874e-004	1.5341e-004	1.8661e-001	1.1550e-005	2.1837e+008	2.9338e+006	2.8578e+007	1.3979e+006

Table 8(b). Third aerobic reactor (Unit 5)

	SN_5		SNH_5	
	INCBP	ARLS	INCBP	ARLS
Computation time for model identification (sec)	7.3788e+000	.8657e+001	8.5801e+000	4.0903e+001
Total square error (TSE)	7.7051e+001	1.9588e+000	2.5831e+000	3.1102e-001
Minimum performance index	2.9074e-001	1.3800e-004	2.0936e-001	1.2437e-004
Mean error of one step ahead prediction of training data	1.1384e+002	8.6449e-001	8.3167e-001	3.2106e-003
Mean error of one step prediction of test data	1.9476e-002	6.6199e-006	1.3913e-001	1.9668e-004
Mean value of 5-step ahead prediction error	1.1841e+001	9.1146e-001	2.0259e+002	1.2591e+002
Akaike's final prediction error (AFPE) estimate	3.4235e+003	1.5950e+000	3.8309e+000	2.2914e-003

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 3.3 and 4.2.3. 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, 5, 6, 7 and 8 for the five reactors. 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, 5, 6, 7 and 8.

The total square error (TSE) discussed in subsection 3.1, for the network trained with the INCBP and the ARLS algorithms are given in the second rows of Table 4, 5, 6, 7 and 8. 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 mean square error (MSE) 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 AS-WWTP process giving smaller errors than the INCBP model.

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

According to the discussion on network validation in Section 3.4, 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 AS-WWTP process.

The results shown in Fig. 7, Fig. 8 and Fig. 9 each having (a) to (e) corresponds to the five reactors respectively. The output parameters obtained in each of (a) to (e) were previously defined during the problem formulation in Section 4.2.1 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 for the anaerobic reactor (Unit 1) for the nitrate concentration SNO_1 in $g.m^{-3}$, the influent flow rate Q_{IN} (Qinf) in $m^3.d^{-1}$, the RAS recycle flow rate Q_{R1} (Q_{ras}) in $m^3.d^{-1}$, and the aeration intensity $K_La1 = 1\ hr^{-1}$.

(b) is for the anoxic reactor (Unit 2) for nitrate and nitrite nitrogen SNO2 in $g.m^{-1}$, particulate products arising from biomass decay XP2 in $g.m^{-1}$, internal recycled nitrate 1 flow rate Q_{A1} (Qirf) in $m^3.d^{-1}$), soluble biodegradable organic nitrogen SND2 in g/m^3 and the aeration intensity K_La2 in hr^{-1}.

(c) is for the first aerobic reactor (Unit 3) for particulate product arising from biomass decay XP3 in $g.m^{-3}$, soluble oxygen SO3 in $g.m^{-3}$ and nitrate and nitrite nitrogen SNO3 in $g.m^{-1}$ and the aeration intensity K_La3 in hr^{-1}.

(d) is for the second aerobic reactor (Unit 4) for readily biodegradable substrate SS4 in $g.m^{-3}$, soluble oxygen SO4 in $g.m^{-3}$, nitrate and nitrite nitrogen SNO4 in $g.m^{-1}$ and the aeration intensity K_La4 in hr^{-}.

(e) is for third aerobic reactor (Unit 5) for ammonia and ammonium nitrogen SNH5 in $g.m^{-3}$, soluble biodegradable organic nitrogen SND5 in $g.m^{-1}$, soluble oxygen SO5 in $g.m^{-3}$, the internal recycled nitrate 2 flow rate Q_{A2} (Qirn) in $m^3.d^{-1}$, feed flow rate Q_{F1} (Qffr) in $m^3.d^{-1}$ and the aeration intensity K_La5 in hr^{-1}.

4.3.1. 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 Table 4, 5, 6, 7 and 8 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 Table 4 to 8. These small one-step ahead prediction errors are indications that the networks trained using the ARLS captures and approximate the nonlinear dynamics of the five reactors 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 Table 4 to Table 8.

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 five reactors 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 Table 4 to Table 8. 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, 5, 6, 7 and 8 verifies the neural network ability to model accurately the dynamics of the five reactors of the AS-WWTP process based on the dry-weather influent data using the proposed ARLS training algorithm.

4.3.2. *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 3.4 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.

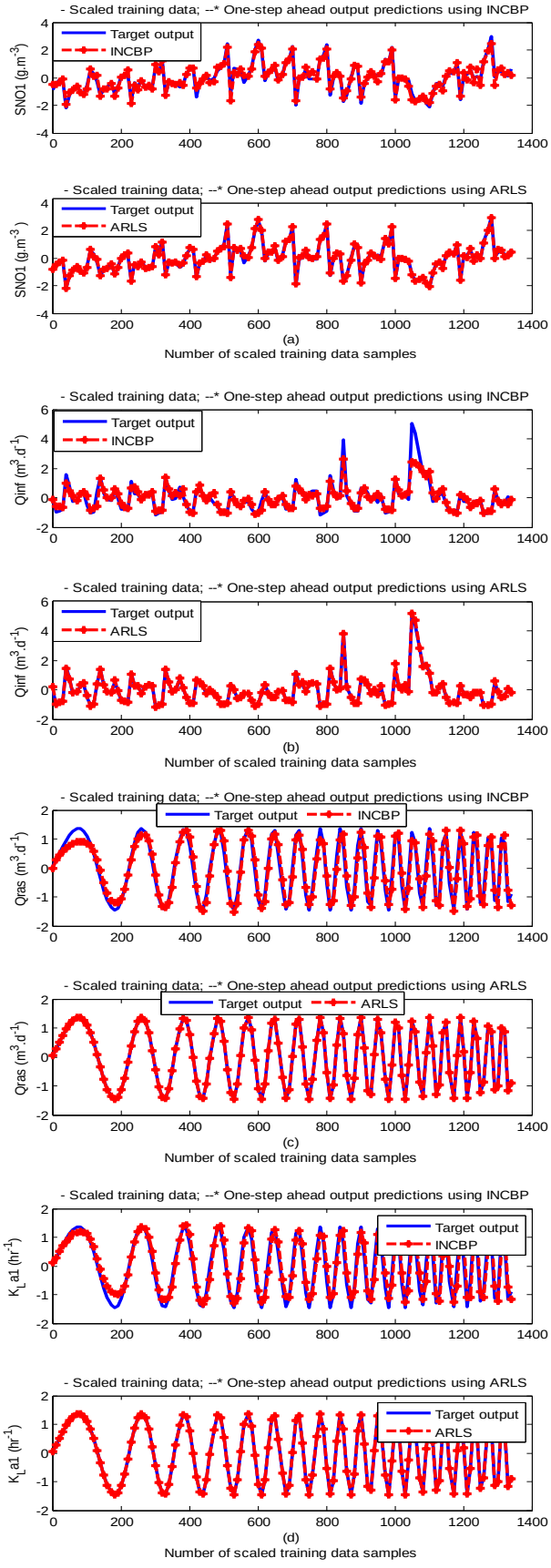


Figure 7(a). One-step ahead prediction of scaled SNO_1, Q_{IN} (Qinf), Q_{R1} (Q_{ras}) and K_La1 training data

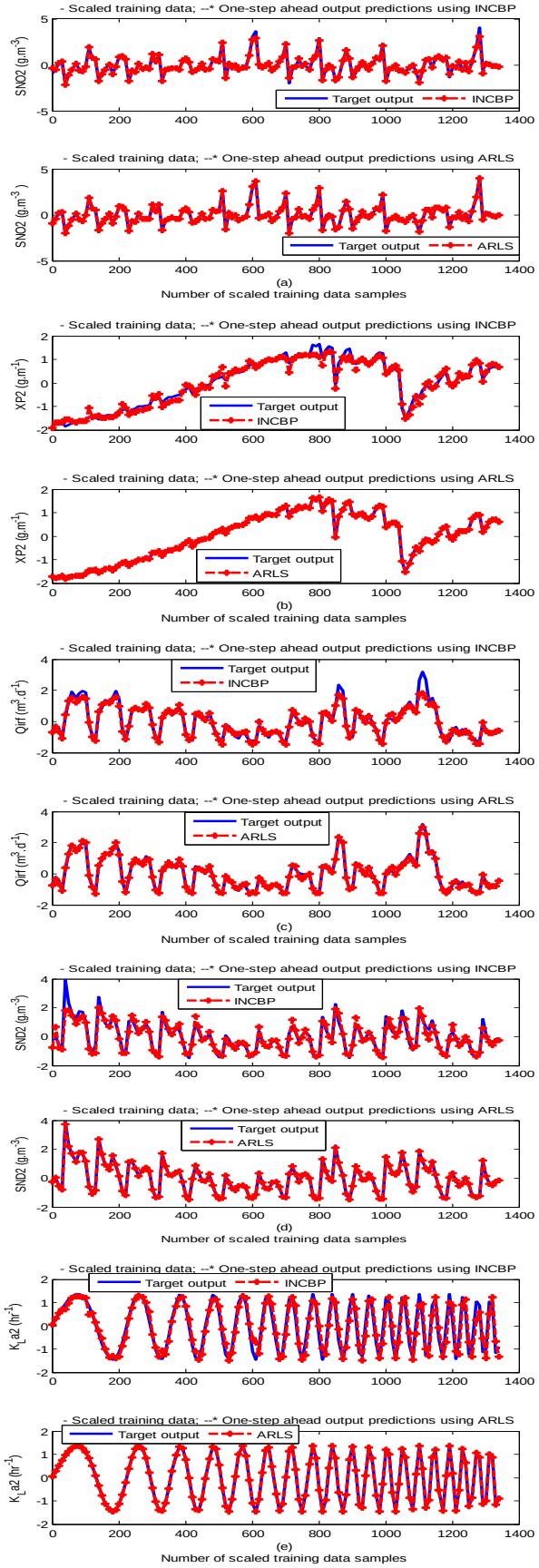


Figure 7(b). One-step ahead prediction of scaled SNO2, XP2, Q_{A1} (Qirf), SND2 and K_La2 training data

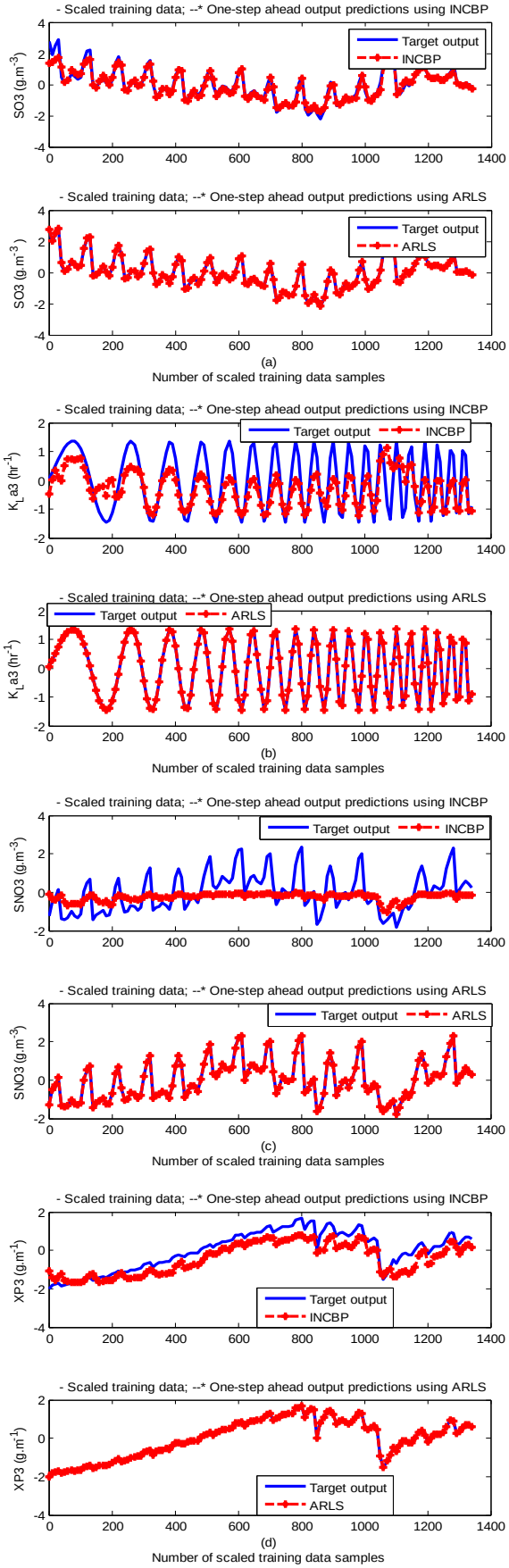


Figure 7(c). One-step ahead prediction of scaled SO3, K_La3, SNO3 and XP3 training data

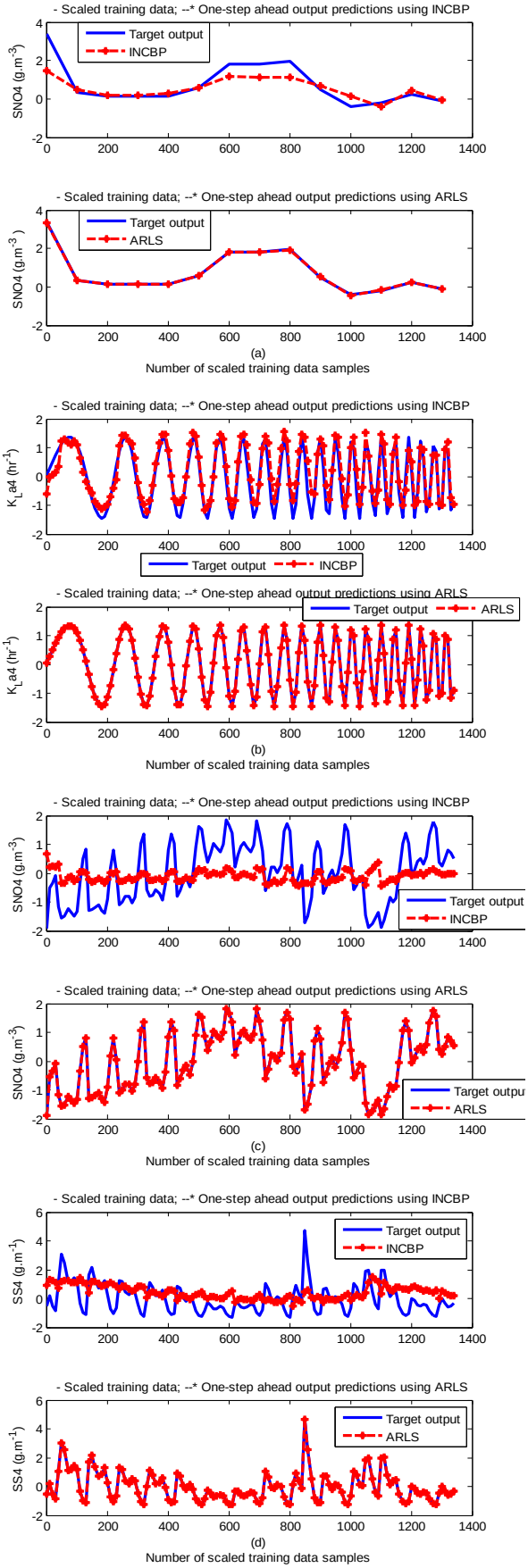


Figure 7(d). One-step ahead prediction of scaled SO4, K_La4, SNO4 and SS4 training data

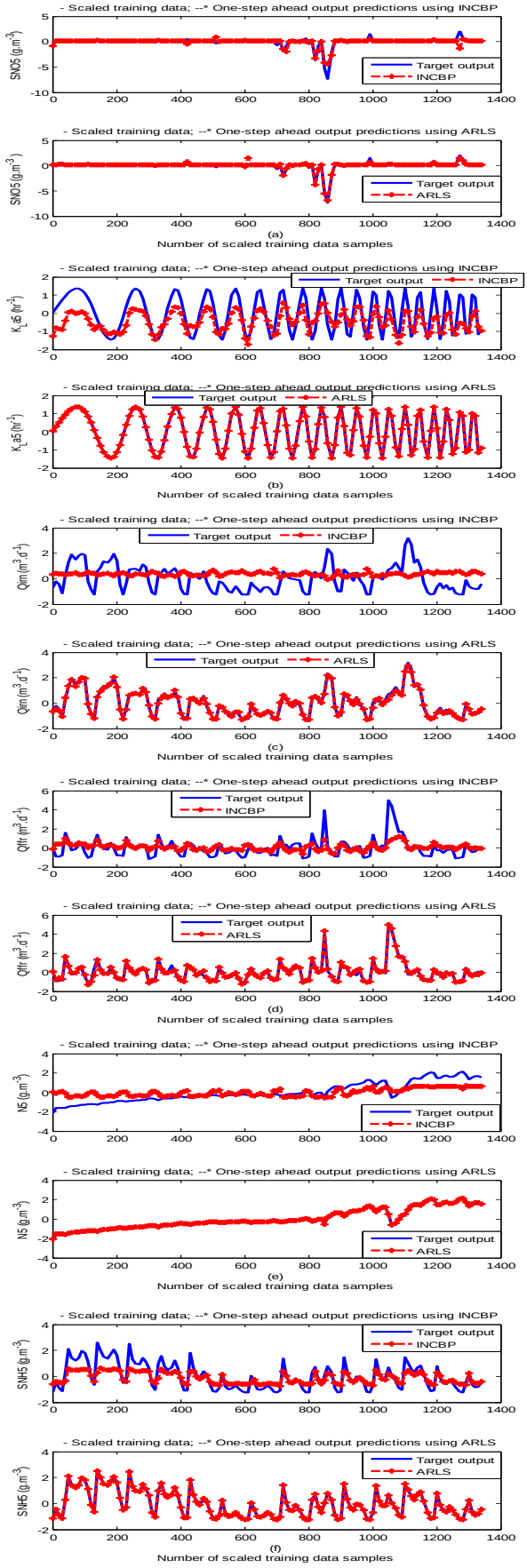


Figure 7(e). One-step ahead prediction of scaled SO5, K_La5, Q_{A2} (Qirn), Q_{F1} (Qffr), SND5 and SNH5 training data

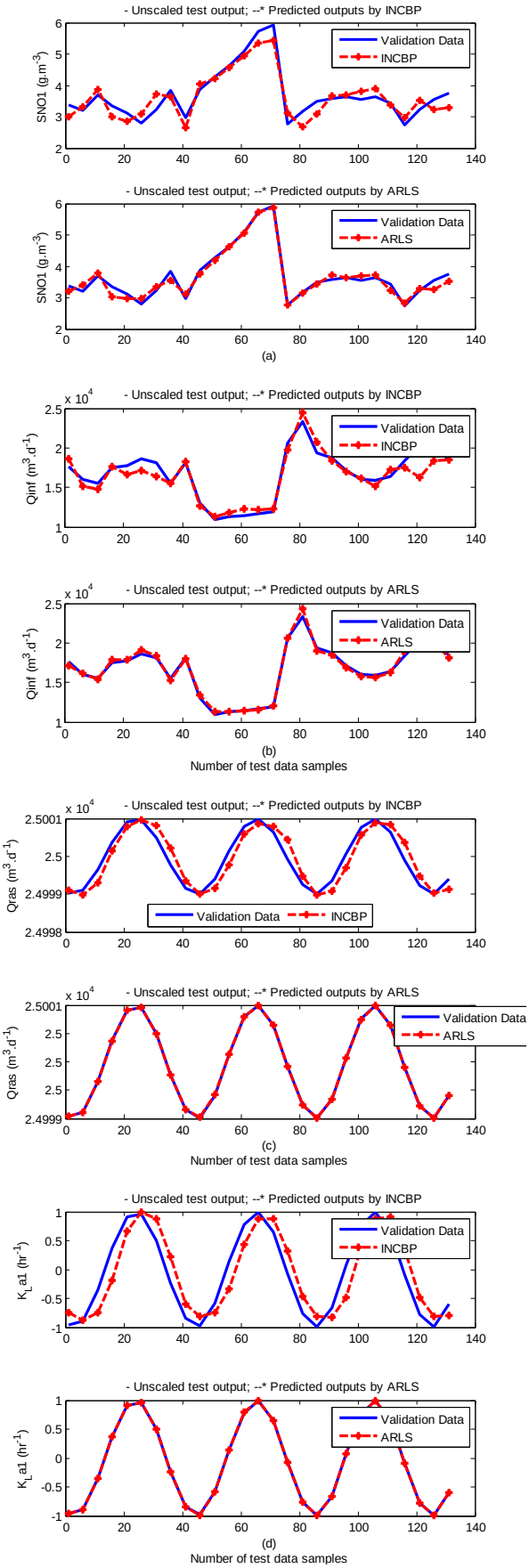


Figure 8(a). One-step ahead prediction of unscaled SNO_1, Q_{IN} (Qinf), Q_{R1} (Q_{ras}) and K_La1 validation data

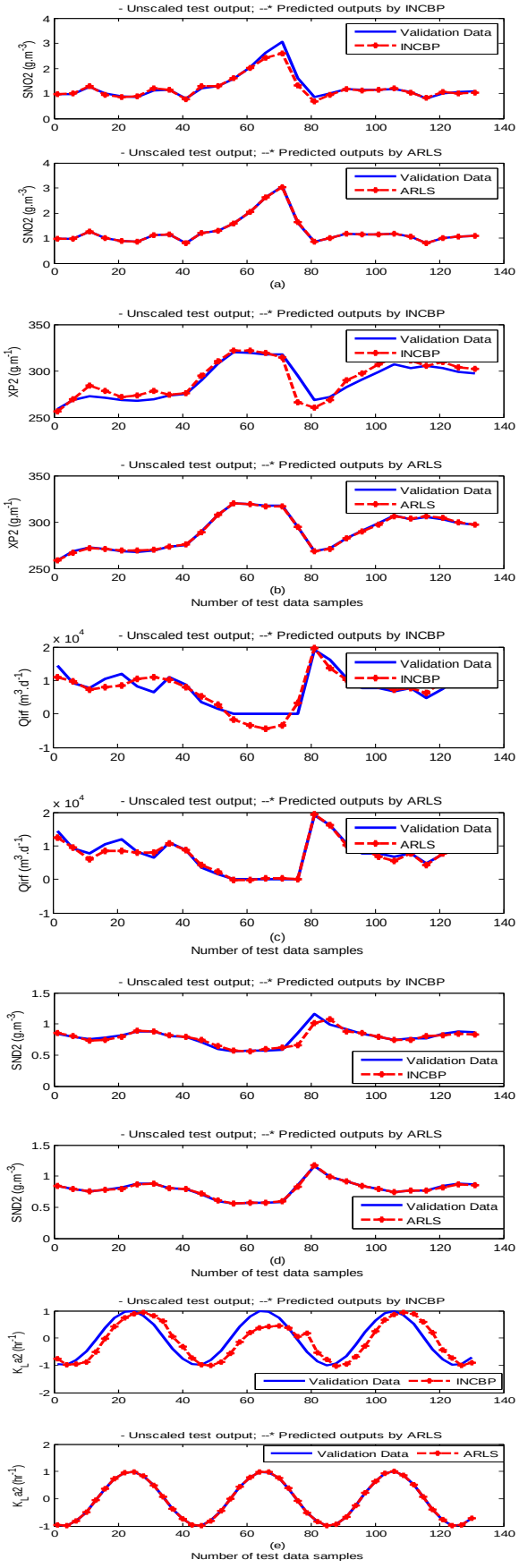


Figure 8(b). One-step ahead prediction of unscaled SNO2, XP2, Q_{A1} (Qirf), SND2 and K_La2 validation data

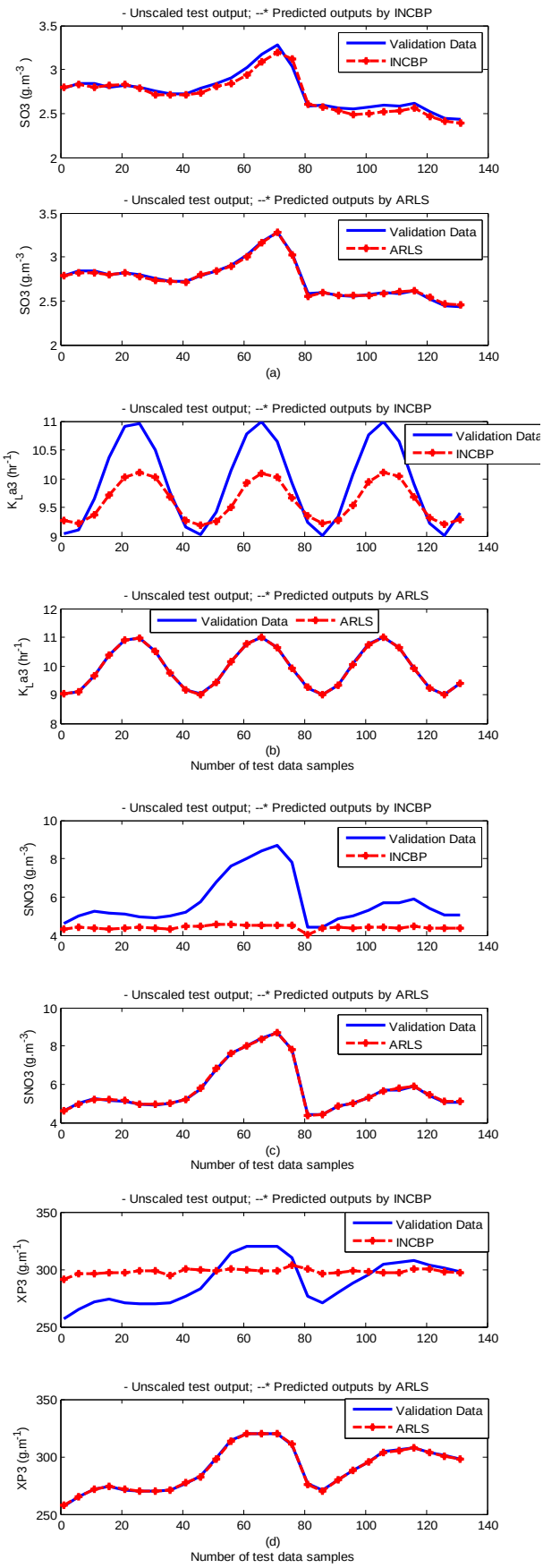


Figure 8(c). One-step ahead prediction of unscaled SO3, K_La3, SNO3 and XP3 validation data

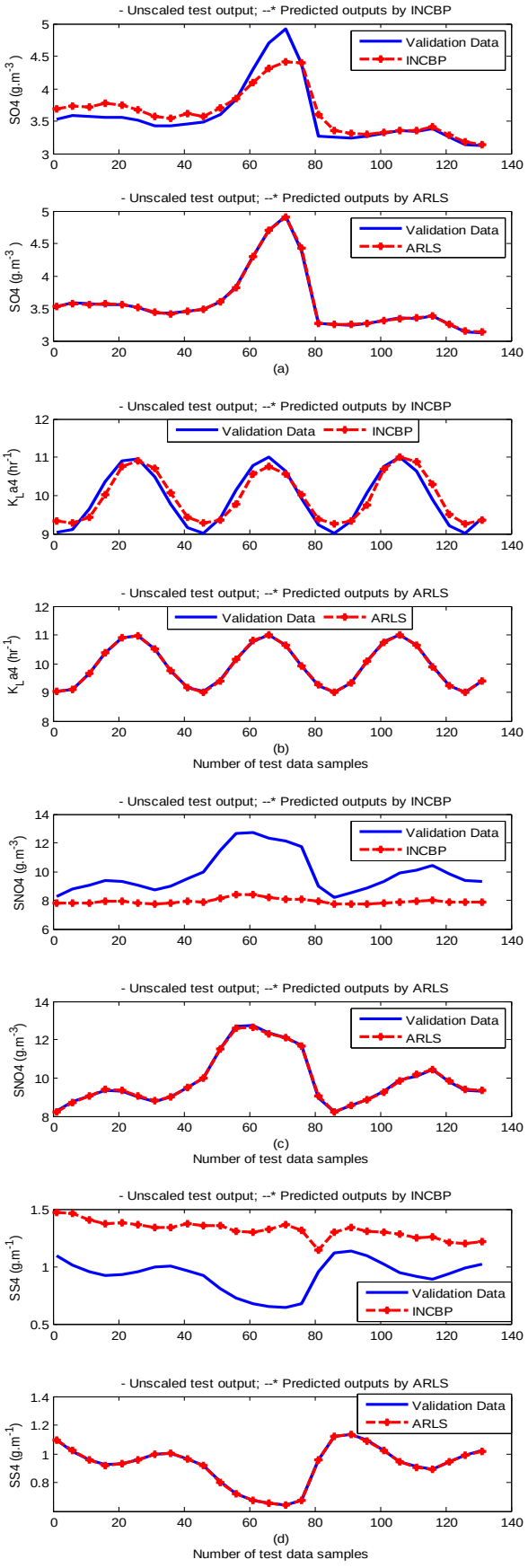


Figure 8(d). One-step ahead prediction of unscaled SO4, K_La4, SNO4 and SS4 validation data

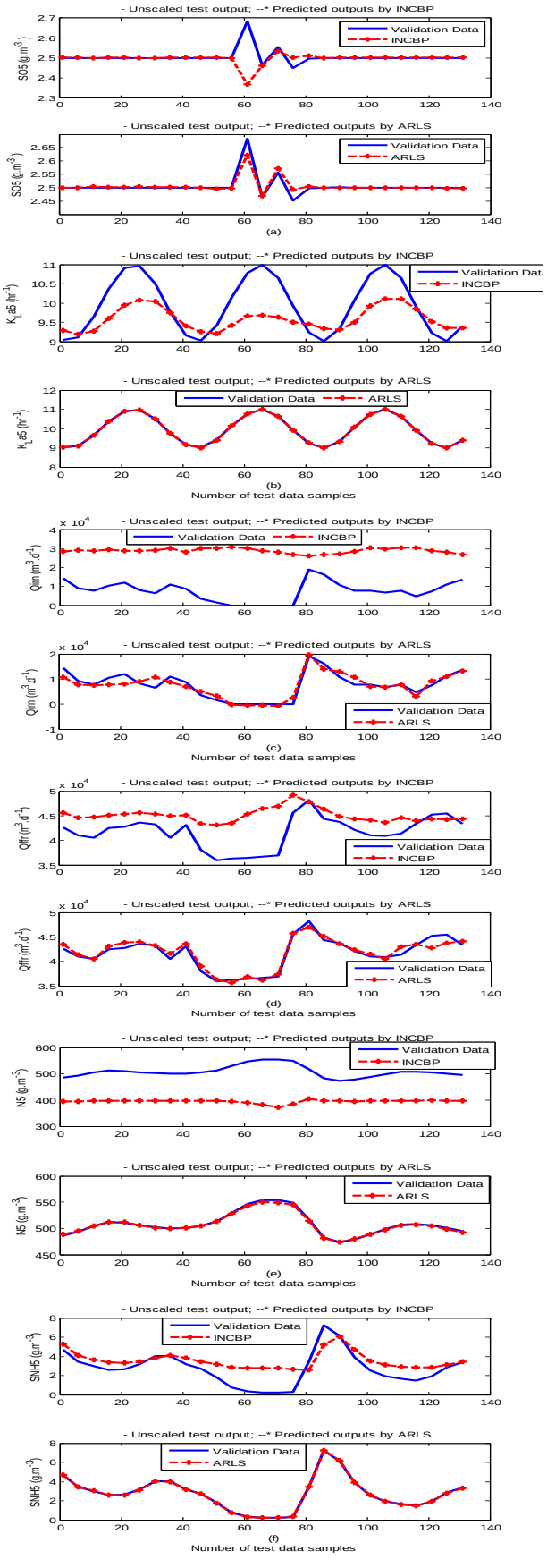


Figure 8(e). One-step ahead prediction of unscaled SO5, K_La5, Q_{A2} (Qirn), Q_{F1} (Qffr), SND5 and SNH5 validation data

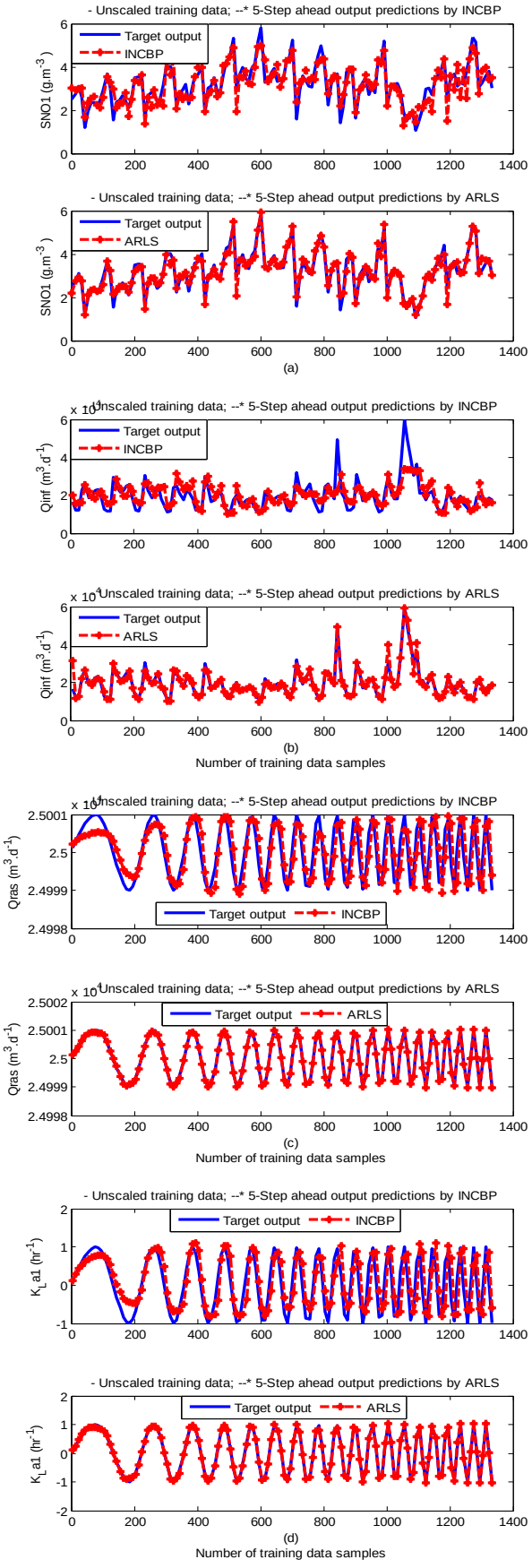


Figure 9(a). 5-step ahead prediction of unscaled SNO_1, Q_{IN} (Qinf), Q_{R1} (Q_{ras}) and K_La1 training data

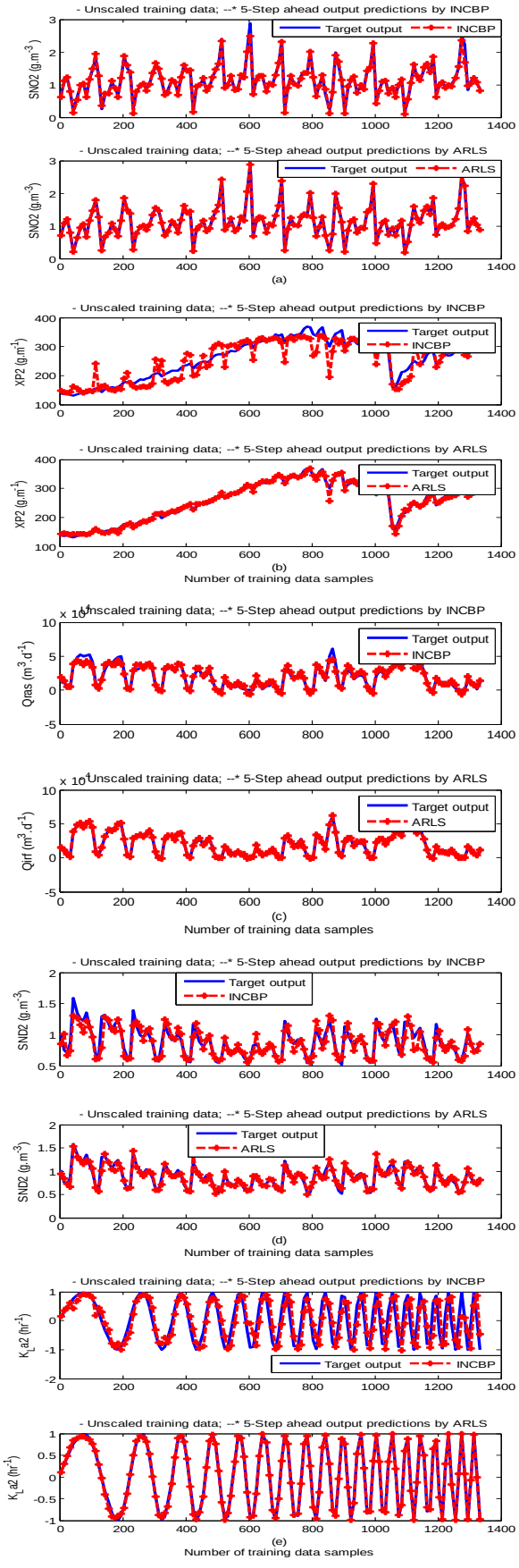


Figure 9(b). 5-step ahead prediction of unscaled SNO2, XP2, Q_{A1} (Qirf), SND2 and K_La2 training data

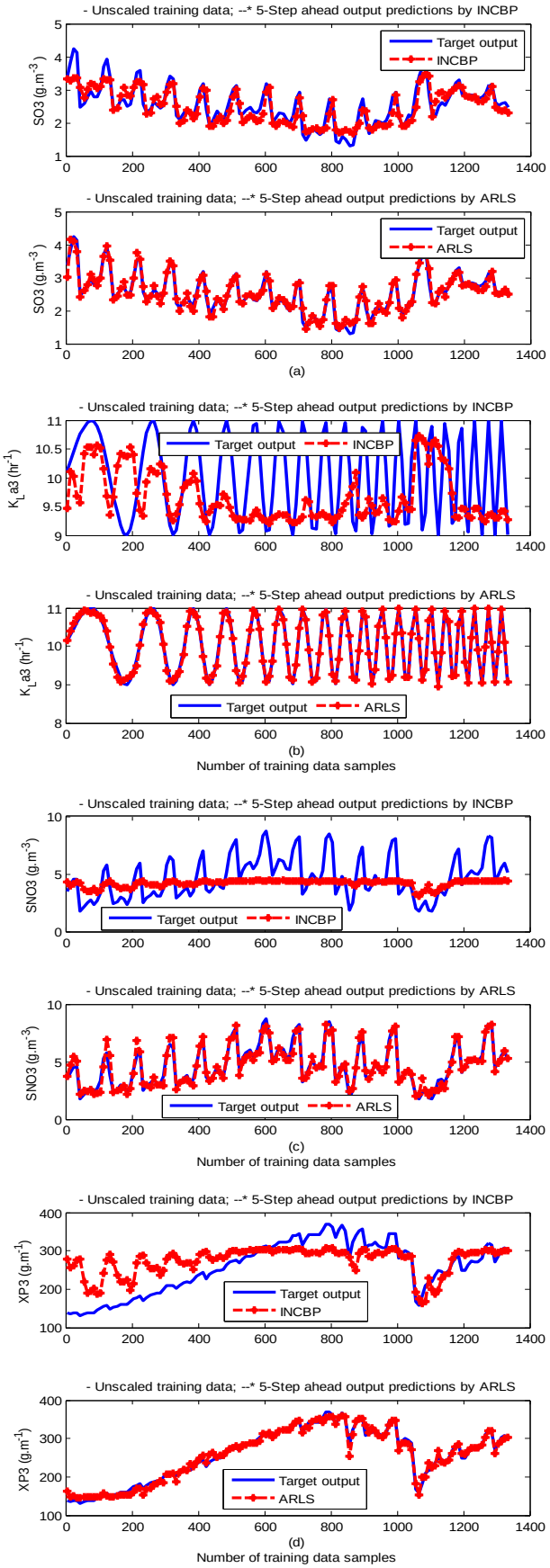


Figure 9(c). 5--step ahead prediction of unscaled SO3, KLa3, SNO3 and XP3training data

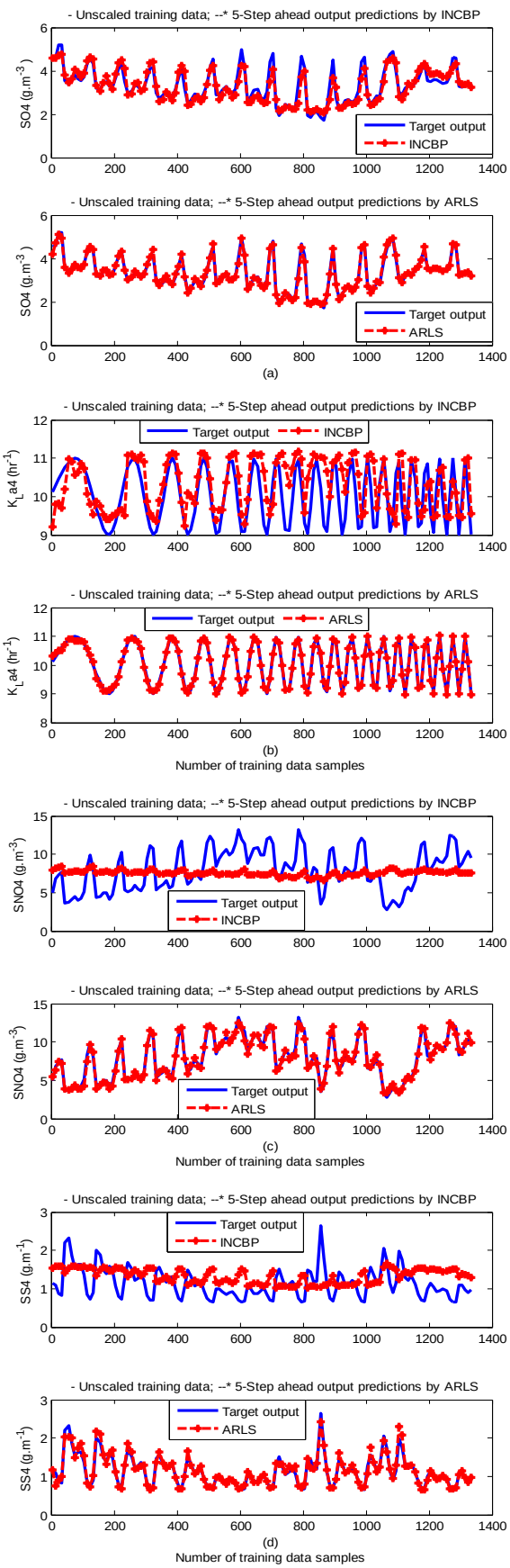


Figure 9(d). 5-step ahead prediction of unscaled SO4, K_La4, SNO4 and SS4 training data

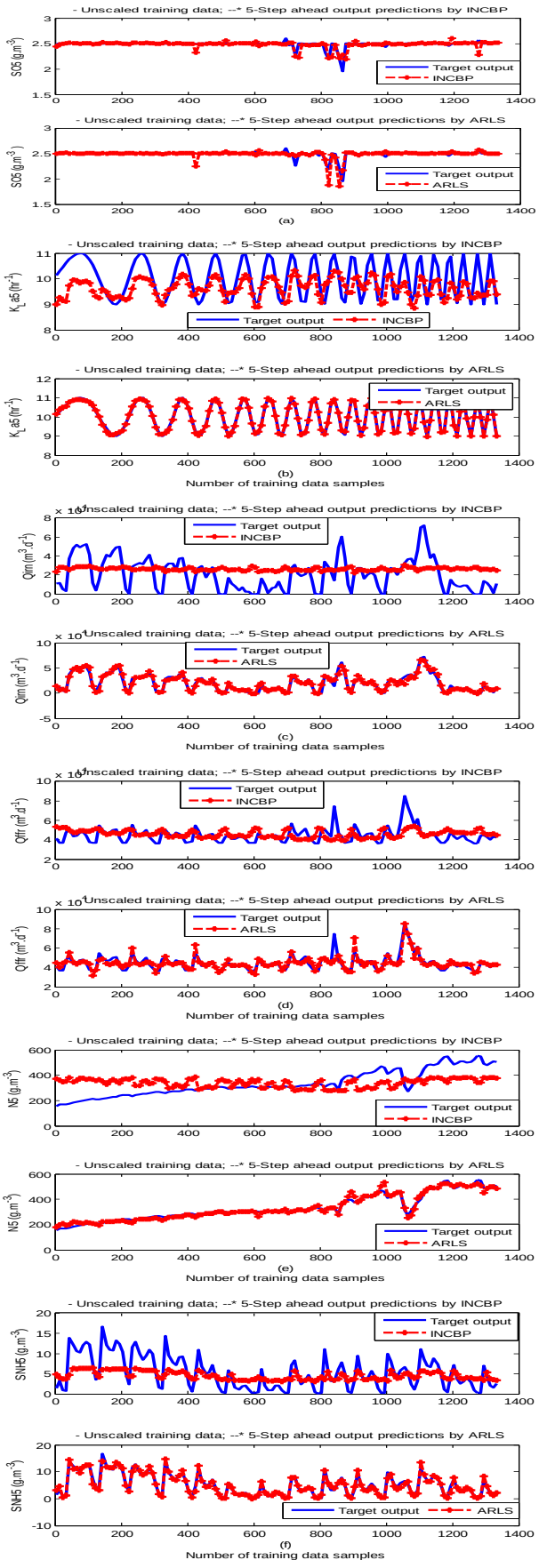


Figure 9(e). 5-step ahead prediction of unscaled SO5, KLa5, Q_{A2} (Qirn), Q_{F1} (Qffr), SND5 and SNH5 training data

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, 5, 6, 7 and 8 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 of the five reactors 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.

4.3.3. Akaike's Final Prediction Error (AFPE) Estimates for the AS-WWTP Process

The implementation of the AFPE algorithm discussed in Section 3.4 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, 5, 6, 7 and 8 respectively. These relatively small values of the AFPE estimate indicate that the trained networks capture the underlying dynamics of the aerobic reactor of the AS-WWTP and that the network is not over-trained [34]. 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.

5. 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* 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 compared with proposed ARLS. The results from the application of these algorithms to the modeling of the five 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 dynamic modelling and nonlinear model identification of the multivariable NNARMAX model of the secondary settler and the clarifier as well as effluent tank to complete the modelling of the

AS-WWTP process. As a future work, the development of an adaptive fuzzy rule-based logic decision system which would be able 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 some decision parameters could be considered.

Appendix

Appendix A: AS-WWTP Process Model

As mentioned in above, the BSM1 model involves eight different chemical reactions (ρ_j) incorporating thirteen different components [1], [27]–[29]. 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. The 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 typical schematic of the AS-WWTP is shown in Fig. 1.

Table A1. Stiochiometric parameters with their units and values

Parameters	Unit	Value
Y_A	g cell COD formed.(g N oxidized)$^{-1}$	0.24
Y_H	g cell COD formed.(g COD oxidized)$^{-1}$	0.67
f_P	Dimensionless	0.08
i_{XB}	g N.(g COD)$^{-1}$ in biomass	0.08
i_{XP}	g N.(g COD)$^{-1}$ in particulate products	0.06

The eight basic processes that are used to describe the biological behaviour of the AS-WWTP process are:

$j=1$: Aerobic growth of heterotrophs

$$\rho_1 = \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_S + S_O} \right) X_{B,H} \quad \text{(A.1)}$$

$j=2$: Anoic growth of heterotrophs

$$\rho_2 = \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \times \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \eta_g X_{B,H} \quad \text{(A.2)}$$

$j=3$: Aerobic growth of autotrophs

$$\rho_3 = \mu_H \left(\frac{S_{NH}}{K_{NH} + S_{NH}} \right) \left(\frac{S_O}{K_{O,A} + S_O} \right) X_{B,A} \quad \text{(A.3)}$$

$j=4$: Decay of heterotrophs

$$\rho_4 = b_H X_{B,H} \quad \text{(A.4)}$$

$j=5$: Decay of autotrophs

$$\rho_5 = b_A X_{B,A} \quad \text{(A.5)}$$

$j=6$: Ammonification of soluble organic nitrogen

$$\rho_6 = k_a S_{ND} X_{B,H} \quad \text{(A.6)}$$

$j=7$: Hydrolysis of entrapped organics

$$\rho_7 = k_h \frac{X_S / X_{B,H}}{K_X + \left(X_S / X_{B,H} \right)} \times \left[\left(\frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right] X_{B,H} \quad \text{(A.7)}$$

$j=8$: Hydrolysis of entrapped organic nitrogen

$$\rho_8 = k_h \frac{X_S / X_{B,H}}{K_X + \left(X_S / X_{B,H} \right)} \times \left[\left(\frac{S_O}{K_{O,H} + S_O} \right) + \eta_h \left(\frac{K_{O,H}}{K_{O,H} + S_O} \right) \left(\frac{S_{NO}}{K_{NO} + S_{NO}} \right) \right] \times \left(X_{B,H} \left(X_{ND} / X_S \right) \right) \quad \text{(A.8)}$$

The observed thirteen conversion rates (r_i) result from combinations of basic processes (A.1) to (A.8) as follows:

$$S_I (i=1): \quad r_1 = 0 \quad \text{(A.9)}$$

$$S_S (i=2): \quad r_2 = -\frac{1}{Y_H} \rho_1 - \frac{1}{Y_H} \rho_2 + \rho \quad \text{(A.10)}$$

$$X_I (i=3): \quad r_3 = 0 \quad \text{(A.11)}$$

$$X_S (i=4): \quad r_4 = (1 - f_P) \rho_4 + (1 - f_P) \rho_5 - \rho_7 \} \quad \text{(A.12)}$$

$$X_{B,H} (i=5): \quad r_5 = \rho_1 + \rho_2 - \rho_4 \quad \text{(A.13)}$$

$$X_{B,A} (i=6): \quad r_6 = \rho_3 - \rho_5 \quad \text{(A.14)}$$

$$X_P (i=7): \quad r_7 = f_P \rho_4 + f_P \rho_5 \quad \text{(A.15)}$$

$$S_O (i=8): \quad r_8 = -\frac{1 - Y_H}{Y_H} \rho_1 - \frac{4.57 - Y_A}{Y_A} \rho_3 \quad \text{(A.16)}$$

Table A2. Kinetic parameters with their units and values

Parameters	Unit	Value
μ_H	$(day)^{-1}$	4.0
K_S	g COD.m^{-3}	10.0
$K_{O,H}$	g (–COD).m^{-3}	0.2
K_{NO}	g NO_{3^-} N.m^{-3}	0.5
b_H	$(day)^{-1}$	0.3
η_g	Dimensionless	0.8
η_h	Dimensionless	0.8
k_h	g slowly biodegradable COD.(g cell COD.$day)^{-1}$	3.0
K_X	g slowly biodegradable COD.(g cell $COD)^{-1}$	0.1
μ_A	$(day)^{-1}$	0.5
K_{NH}	g NH_{3^-} N.m^{-3}	1.0
b_A	$(day)^{-1}$	0.05
$K_{O,A}$	G (COD).m^{-3}	0.4
k_a	m^{-3}.(g COD.$day)^{-1}$	0.05

$$S_{NO}(i=9):\quad r_9 = -\frac{1-Y_H}{2.86Y_H}\rho_2 + \frac{1}{Y_A}\rho_3 \qquad \text{(A.17)}$$

$$S_{NH}(i=10):\quad \left.\begin{aligned} r_{10} &= -i_{XB}\rho_1 - i_{XB}\rho_2 \\ &\quad -\left(i_{XB}+\frac{1}{Y_A}\rho_3\right)+\rho_6 \end{aligned}\right\} \qquad \text{(A.18)}$$

$$S_{ND}(i=11):\qquad r_{11} = -\rho_6 + \rho_8 \qquad \text{(A.19)}$$

$$X_{ND}(i=12):\quad \left.\begin{aligned} r_{12} &= \left(i_{XB} - f_P i_{XP}\right)\rho_4 \\ &\quad +\left(i_{XB} - f_P i_{XP}\right)\rho_5 - \rho_8 \end{aligned}\right\} \qquad \text{(A.20)}$$

$$S_{ALK}(i=13):\quad \left.\begin{aligned} r_{13} &= -\frac{i_{XB}}{14}\rho_1 + \left(\frac{1-Y_H}{14\times 2.86Y_H} - \frac{i_{XB}}{14}\right)\rho_2 \\ &\quad -\left(\frac{i_{XB}}{14}+\frac{1}{7Y_A}\right)\rho_3 + \frac{1}{14}\rho_6 \end{aligned}\right\} \qquad \text{(A.21)}$$

The biological parameter values used in the BSM1 correspond approximately to a temperature of 15 ℃. The stiochiometric parameters are listed in Table A.1 and the kinetic parameters are listed in Table A.2.

Appendix B: General Characteristics of the Biological Reactors

As shown in Fig. 1, the general characteristics of the biological reactors for the default case are five compartments where the first two (Unit 1 and Unit 2) are non-aerated compartments whereas the last three (Unit 3, Unit 4 and Unit 5) are aerated compartments.

Unit 3 and Unit 4 of the aerated compartments have a fixed oxygen transfer coefficient of $K_L a = 10\,h^{-1} = 240\,day^{-1}$. In Unit 5, the dissolved oxygen (DO) concentration is controlled at a level of $2\,g\,(-COD).m^{-3}$ by manipulation of the $K_L a$. Each of the five compartments has a flow rate Q_k, the concentration Z_k, and the reaction rate r_k; where $k = 1, 2, \ldots, 5$ is the number of compartments. The volume of the non-aerated compartments is $1{,}000\,m^3$ each while the volume of the aerated compartments is $1{,}333\,m^3$.

The general equation for the reactor mass balances is given as:

For $k = 1$ (Unit 1)

$$\frac{dZ_1}{dt} = \frac{1}{V_1}\left(Q_a Z_a + Q_r Z_r + Q_0 Z_0 - Q_1 Z_1 + r_1 V_1\right) \qquad \text{(B.1)}$$

$$Q_1 = Q_a + Q_r + Q_0 \qquad \text{(B.2)}$$

For $k = 2$ to 5 (Unit 2, Unit3, Unit 4 and Unit 5)

$$\frac{dZ_k}{dt} = \frac{1}{V_k}\left(Q_{k-1}S_{O,k-1} - Q_k Z_k + r_k V_k\right) \qquad \text{(B.3)}$$

$$Q_k = Q_{k-1} \qquad \text{(B.4)}$$

Special case for oxygen $(S_{O,k})$:

$$\left.\begin{aligned} \frac{dZ_k}{dt} &= \frac{1}{V_k}\left(Q_{k-1}S_{O,k-1} + r_k V_k\right. \\ &\quad \left. +\left(K_L a\right)_k V_k\left(S_O^* - S_{O,k}\right) - Q_k S_{O,k}\right) \end{aligned}\right\} \qquad \text{(B.5)}$$

where the saturation concentration for oxygen is $S_O^* = 0.8\,g.m^{-3}$. Also,

$$Z_a = Z_5 \tag{B.6}$$

$$Z_f = Z_5 \tag{B.7}$$

$$Z_w = Z_r \tag{B.8}$$

$$\left.\begin{aligned} Q_f &= Q_5 - Q_a = Q_e + Q_r + Q_w \\ &= Q_e + Q_u \end{aligned}\right\} \tag{B.9}$$

REFERENCES

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

[2] K. R. Muske and J. B. Rawlings, "Model predictive control with linear models," *AIChE Journal*, vol. 39, pp. 262-287, 1993.

[3] L. Kalra and C. Georgakis, "Effects of process nonlinearity on the performance of linear model predictive controllers for the environmentally safe operation of a fluid catalytic cracking unit," *Ind. Eng. Chem. Res*, vol. 33, pp.3063-3069, 1994.

[4] I. Dones, F. Manenti, H. A. Preisig and G. Buzzi-Ferraris, "Nonlinear model predictive control: A self-adaptive approach". *Ind. Eng. Chem. Res.*, vol. 49, no. 10, pp. 4782 – 4791, 2010.

[5] P. Potočnik and I. Grabec, "Nonlinear model predictive control of a cutting process". *Neurocomputing*, vol. 43, pp. 107 – 126, 2002.

[6] C. Su and Y. Wu, "Adaptive neural network predictive control based on PSO algorithm," *Chinese Control and Decision Conference*, Guilin, China, June 17-19, pp. 5829 – 5833, 2009.

[7] E. F. Camacho and C. Bordons, "Model Predictive Control". 2nd ed., *London: Springer-Verlag*, 2007.

[8] M. J. Grimble and A. W. Ordys, "Predictive control for industrial applications". *Annual Reviews in Control*, vol. 25, pp. 13-24, 2001.

[9] J. M. Maciejowski, "Predictive Control with Constraints". *England: Pearson Education Limited*, 2002.

[10] Y. Jin and C. Su, "Adaptive model predictive control using diagonal recurrent neural network". *Fourth Int'l. Conf. on Natural Computation*, Jinan, Oct. 18-20, 2008, pp. 276 – 280, 2008.

[11] F. S. Mjalli, "Adaptive and predictive control of liquid-liquid extractors using neural-based instantaneous linearization technique". *Chem. Eng. Technol.*, vol. 29, no. 5, pp. 539 – 549, 2006.

[12] 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.

[13] 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.

[14] Omidvar, O. M. and Elliott, D. L. (Feb., 1997) "Neural systems for control". *Academic Press, San Diego*. [Online] Available: http://www.isr.umd.edu/~delliot/NeuralSystemsForControl.pdf.

[15] K. Salahshoor, E. Safari and M. F. Samadi, "Adaptive model predictive control of a hybrid motorboat using self-organizing GAP-RBF neural network and GA algorithm". *2nd IEEE Int'l Conf. on Adv. Computer Control*, Shenyang, China, Mar. 27-29, 2010, pp. 588 – 592, 2010.

[16] J. Sarangapani, "Neural Network Control of Discrete-Time Systems". *Boca Raton: CRC Press*, 2006.

[17] T. J. Spooner, M. Maggiore, R. Ordóñez and K. M. Passino, "Stable Adaptive Control and Estimation for Nonlinear Systems: Neural and Fuzzy Approximator Techniques". *New York: John Wiley & Sons*, 2002.

[18] G. I. Suárez, O. A. Ortiz, P. M. Aballay and N. H. Aros, "Adaptive neural model predictive control for the grape juice concentration process," *2010 IEEE Int'l Conf. on Industrial Tech.*, Vi a del Mar, Mar. 14-17, pp. 57 – 63, 2010.

[19] D. W. Yu and D. L. Yu, "Multi-rate model predictive control of a chemical reactor based on three neural models". *Biochemical Engineering Journal*, vol. 37, pp. 86 – 97, 2007.

[20] P. Guarneri, G. Rocca and M. Gobbi, "A neural network-based model for the dynamic simulation of the tire/suspension system while traversing road irregularities". *IEEE Trans. Neural Netw.*, vol.19, no. 9, pp. 1549 – 1563, 2008.

[21] U. Yüzgeç, Y. Becerikli, and M. Türker, "Dynamic neural-network-based model predictive control of an industrial baker's yeast drying process". *IEEE Trans. Neural Networks*, vol. 19, no. 7, pp. 1231 – 1242, 2008.

[22] C. H. Lu and C. C. Tsai, "Adaptive predictive control with recurrent neural network for industrial process: An application to temperature control of a variable-frequency oil-cooling machine". *IEEE Trans. Industrial Electronics*, vol. 55, no. 3, pp. 1366 – 1375, 2008.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

[29] 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.

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

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

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

[33] K.S. Narendra and O.A. Driollet, Stochastic adaptive control using multiple estimation models. In *Proc. of the 2001 American Control Conference (ACC2001)*, Arlington, VA, 25 – 27 June, 2001, pp. 1539 – 1544.

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

[35] 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.

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

[37] R. Fletcher, *Practical Methods of Optimization.* 2nd ed. Wiley & Sons, 1987.

[38] J. Hertz, A. Krough and R. G. Palmer, "*Introduction to the Theory of Neural Computing*", Redwood City, California: Addison-Wesley, 1991.