

Article

Not peer-reviewed version

LSTM Network for the Oxygen Concentration Modeling of a Wastewater Treatment Plant

[Chiara Toffanin](#)*, Federico Palma, [Francesca Iacono](#), Lalo Magni

Posted Date: 8 May 2023

doi: 10.20944/preprints202305.0498.v1

Keywords: Black-box models; Neural networks; LSTM; oxygen concentration modeling



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

LSTM Network for the Oxygen Concentration Modeling of a Wastewater Treatment Plant

Chiara Toffanin ^{1,*}, Federico Di Palma ², Francesca Iacono ¹ and Lalo Magni ²

¹ Department of Electrical, Computer and Biomedical Engineering, University of Pavia, via Ferrata 3, Pavia, 27100, Italy; chiara.toffanin@unipv.it, francesca.iacono01@universitadipavia.it

² Department of Civil and Architecture Engineering, University of Pavia, via Ferrata 3, Pavia, 27100, Italy; federico.dipalma@unipv.it, lalo.magni@unipv.it

* Correspondence: chiara.toffanin@unipv.it

Abstract: Activated sludge process is a well known method to treat municipal and industrial waste water. In this complex process, the oxygen concentration in the reactors plays a critical role in the plant efficiency. This paper proposes the use of a Long Short-Term Memory (LSTM) network to identify an input-output model suitable for the design of an oxygen concentration controller. The model is identified from easy-accessible measures collected from a real plant. This dataset covers almost a month. The performances achieved with the proposed LSTM model are compared with the ones obtained with a standard AutoRegressive model with exogenous input (ARX). Both models catch the oscillation and the overall behaviour (ARX $\rho = 0.833$, LSTM $\rho = 0.921$), but, while the ARX model fails in reaching the correct amplitude ($FIT = 41.20\%$), the LSTM presents satisfactory performance ($FIT = 60.56\%$).

Keywords: black-box models; neural networks; LSTM; oxygen concentration modeling

1. Introduction

A WasteWater Treatment Plant (WWTP) is a plant where the wastewater is treated to remove pollutants, exploiting biological and chemical reactions before being released back to the environment. The modeling and control of this process is not trivial, since the treatment depends on the nature and the characteristics of the wastewater. Wastewater usually firstly undergoes to chemical-physical treatments, in order to remove the solid part of the wastes, and then to biological treatments for the organic components. One of the most common biological treatment used in WWTPs is the Conventional Activated Sludge (CAS) process, where bacteria are used to nitrify and denitrify wastewater.

Considering the complexity of the process, the control of a WWTP can be a challenging task: several flows may be involved, among with a huge variety of different biological and chemical reactions; it is necessary to provide a sufficient oxygen quantity but without an excess of aeration, reaching a trade-off between energy consumption and process demand, with a growing attention to the environmental related problems. To design an optimal control strategy it is necessary to have an accurate model of the process [1]. In [2] a review on the current state of the art regarding modeling of activated sludge WWTPs is presented.

The model of the entire WWTP is usually composed by two main components: the hydraulic model that takes into account the different flows in input and output in the reactor, and the Activated Sludge Model (ASM) that models the biological and chemical reactions inside the tank due to the activated sludge process. The so called Activated Sludge Model No. 1 (ASM1) [3] was extended in the years to consider the phosphorus dynamics in ASM2 [4] and including storage of organic substrates as a new process with an easier calibration in ASM3 [5]. Anyway, ASM1 is still the most used model and it can be considered the state of the art since several successive works are based on it [6–8]. However, complex and specialized models require the identification of more specific parameters subject to high sensitivity, with expensive and time consuming procedures, far from practical applications. Hence, these specialized models are not a good choice for process design and practical applications and in [9]

the combination of the process knowledge with new artificial intelligence techniques is proposed. Also in [2], in addition to the most diffused WWTP white-box modeling, some black-box methodologies based on Neural Network (NN) techniques were presented with interesting results.

In view of all these considerations, in this work a NN approach is explored to model a WWTP located in Mortara (Pavia, Italy), managed by the company ASMortara S.p.A.. The main goal was the development of a model for the biological reactor of the plant to improve several critical aspects of the process such as the oxygen concentration regulation. The data used in this paper, in fact, have been collected with a simple switching controller enabling (disabling) the input flow if the oxygen concentration inside the reactor is below (above) a certain threshold. The effect of this control strategy is an on/off oxygen flow rate, that generates important oscillations in the oxygen concentration evolution.

2. Materials and Methods

This work is focused on modeling the dynamic of the oxygen in the reactor. In the physical models ASM1, ASM2, ASM3, the oxygen concentration (output), $O(t)$, is mainly connected to few manipulable variables (input): the inlet flow of the wastewater, $q_{in}(t)$, and its Chemical Oxygen Demand, $COD_{in}(t)$; the oxygen inlet flow, $q_{oss}(t)$; the outlet flow of the clarified liquid filtered by the membrane, $q_{ol}(t)$ and the one manually removed, $q_{om}(t)$.

Even if this notion is general, as clearly stated in the Introduction, every plant has its own peculiarity due to the local law limit, the design constraint, the average chemical composition of the waste and other factors [10]. This fact should require a model identification that is typically impossible in view of the limited number of variables measurable on a real plant. Black-box models, and in particular NN, can be a reasonable compromise to overcome this limitation.

In the following, the plant in hand is described in details; then, the proposed Long Short-Term Memory (LSTM) is reported and compared with a classic AutoRegressive model with exogenous input (ARX).

2.1. Plant description

The considered WWTP is an industrial plant actually used to treat both municipal and industrial wastewater. The plant is composed by two different units (see Figure 1) dedicated to the two stages of the treatment, called primary and secondary.

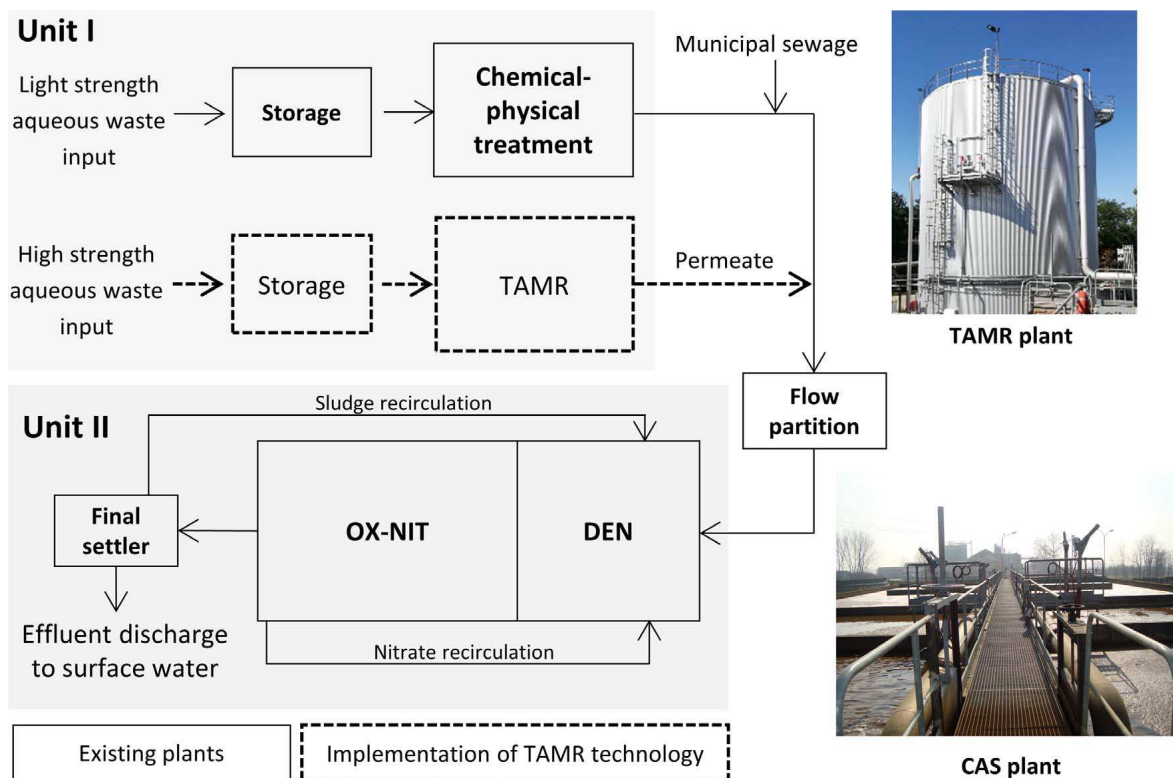


Figure 1. Flow diagram of the WWTP [11].

The industrial wastewater needs to be pre-treated before being accepted in municipal WWTP: it passes in Unit I to remove the solid part of the waste, exploiting chemical reactions. Firstly, the flow goes into a storage-equalization tank, with the goal to transform a variable flow into a steady-state one that goes in input to the treatment plant, optimizing in this way the process. The flow is then divided in two chemical-physical treatment lines: one dedicated to the light strength aqueous wastes and the second for the high strength ones. The light strength wastes are subjected to a chemical-physical treatment with classic procedures like flocculation, coagulation and sedimentation. The wastes are put in sedimentation tanks where the solid part is separated by gravity; chemicals can be added to help coagulation, while with flocculation the small colloidal particles are separated from the wastes and settle in form of flocks. The high strength aqueous waste instead are fed into a Thermophilic Aerobic Membrane Reactor (TAMR). The plant has been recently equipped with this innovative reactor, built ad-hoc to treat high strength wastes. The advantages of the TAMR technology is the exploiting of biological reactions even during the pre-treating phase, considering that biological solutions are more affordable and sustainable of chemical ones. In practice, the waste undergoes to aerobic reactions in thermophilic conditions, exploiting oxygen with temperatures greater than 45 °C [12,13].

After this phase, the pre-treated industrial wastewater is mixed with the municipal wastewater and flow into the biological reactor (Unit II), where the CAS comprehends denitrification (DEN), oxidation and nitrification (OX-NIT). This paper is focus on the modelling of this part of the plant. The biological reactor is filled with oxygen, where aerobic microorganisms are introduced to react with wastewater and to reduce its organic compounds. The overall flow goes to a settling tank (Final settler) where the activated sludge settles, since the micro-organisms create a biological flock, producing a liquid mostly free from suspended solid. The sludge is separated from the clarified water: the majority of the sludge is reintroduced in the CAS to treat the new wastewater, the remaining part is removed, while the cleared water is released in the environment. In the biological reactor the wastewater is subjected also to the denitrification and nitrification processes to remove the nitrogen components from the wastewater. This is done before releasing back the water in the environment since these components are dangerous for aquatic organisms and plants. The nitrification occurs in

aerobic conditions and is the oxidation of ammonia into nitrite and nitrate by nitrifying autotrophic bacteria; the denitrification instead happens in anaerobic conditions to reduce nitrite to nitrogen [14]. For this reason nitrate recirculates from OX-NIT to DEN (Figure 1).

2.2. Long Short-Term Memory (LSTM) network

The LSTM [15,16] is a particular type of Recurrent Neural Network (RNN), that thanks to its specific units architecture, called memory cells, is able to learn the information dependencies better than a simpler RNN. The LSTM memory cell presents an internal self-loop, in addition to the outer feedback present in all the RNNs, and a gating system that regulates the flow of information. In this way, the gradient flows during time, even for long periods, and its derivative do not explode nor vanish.

The mathematical formulation of a single memory cell is represented in Figure 2 and can be expressed as follows:

$$c_k = f_k \times c_{k-1} + i_k \times \tilde{c}_k \quad (1a)$$

$$h_k = o_k \times \tanh(c_k) \quad (1b)$$

where $c_k \in \mathbb{R}^n$ is the internal cell state at time k , $u_k \in \mathbb{R}^m$ is the input at time k and $h_k \in \mathbb{R}^n$ is the hidden state at time k , with n the cell state dimension, that corresponds to the number of LSTM neurons, m the number of features/input of the LSTM, \times is the Hadamard product and \tanh is the hyperbolic tangent, used as activation function.

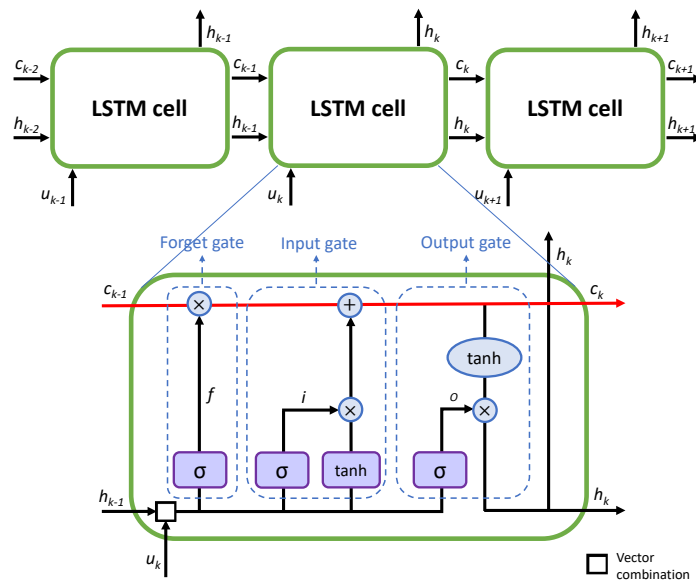


Figure 2. Scheme of a single LSTM cell.

Four structures, called gates, regulates the internal loop of the cell adding or removing information to modify the value of the cell state at each time instant. The first gate is the forget gate, f_k , that decides which information is removed from the cell state c_k , multiplying it by its past value c_{k-1} , as can be seen in the first part of (1a). Then the input gate, i_k , chooses which value of the cell state is updated, while the input activation gate, \tilde{c}_k , creates a new candidate value of the cell state. The last gate is the output gate, o_k , that determines which information contained in the cell state is going in output. The

hidden state is then defined in (1b), where the output gate is multiplied by a transformation of the cell state through hyperbolic tangent function. These gates are expressed as:

$$f_k = \sigma(W_f u_k + U_f h_{k-1} + b_f) \quad (2a)$$

$$i_k = \sigma(W_i u_k + U_i h_{k-1} + b_i) \quad (2b)$$

$$\tilde{c}_k = \tanh(W_{\tilde{c}} u_k + U_{\tilde{c}} h_{k-1} + b_{\tilde{c}}) \quad (2c)$$

$$o_k = \sigma(W_o u_k + U_o h_{k-1} + b_o) \quad (2d)$$

where $W_f, W_i, W_o, W_{\tilde{c}} \in \mathbb{R}^{n \times m}$ are the input weight matrices, $U_f, U_i, U_o, U_{\tilde{c}} \in \mathbb{R}^{n \times n}$ are the previous output weight matrices and $b_f, b_i, b_o, b_{\tilde{c}} \in \mathbb{R}^{n \times 1}$ are the biases. These quantities are specific of each gate and identified during the training of the network. The chosen activation functions are the hyperbolic tangent \tanh and σ , the sigmoid function, equal to:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

Several LSTM cells are combined in the so-called LSTM layer and the LSTM network can be composed by one or more stacked layers, representing the hidden layers. In regression applications, usually, the output of the last (if more) LSTM layer is passed to a fully connected layer without activation function, that computes a linear combination, according to the following equation:

$$y_k = W_y h_k + b_y \quad (4)$$

where $W_y \in \mathbb{R}^{p \times n}$ is the matrix of the output weights, with p the number of outputs of the network, and $b_y \in \mathbb{R}^{p \times 1}$ the bias. W_y and b_y are determined during the training of the network as well.

Thanks to this structure, LSTMs show their strength especially when dealing with temporal data and when temporal lags must be taken into account [16–20]. The LSTM developed in this work considers 3 inputs ($m = 3$): the ingoing oxygen flow rate $q_{oss}(t)$, the ingoing substrate flow rate $q_{in}(t)$ and the outgoing liquid substrate flow rate $q_{ol}(t)$, and one output ($p = 1$), the oxygen concentration $O(t)$. In this case the other inputs of the physical process $q_{om}(t)$ and $COD_{in}(t)$ are not considered in input since they are daily measurements, with very little variations, that can be considered almost constant signals not significantly contributing to the model. All the signals have been rescaled using mean normalization before being used to train the LSTM network.

2.3. AutoRegressive model with eXogenous input (ARX) Model

The ARX model family is the most used model structure to derive input-output models. According to this family [21], the process output y is given by three component: the autoregressive component (the previous value of the output), the exogenous input variable, u , and a white Gaussian additive noise, e . The process output can be defined as

$$y_k = \sum_{j=1}^{N_a} \alpha_j y_{k-j} + \sum_{j=1}^{N_b} \beta_j u_{k-d-j} + e_k$$

where the suffix k indicates the time k of the sample, N_a and N_b are the order of the autoregressives and of the input component respectively, and d is the input delay; those values as a whole are commonly addressed as the *structure* of the ARX model. In this scenario the one step predictor is given by the following

$$\hat{y}_k^p = \sum_{j=1}^{N_a} \alpha_j y_{k-j} + \sum_{j=1}^{N_b} \beta_j u_{k-d-j}$$

where the prediction \hat{y}_k^p is obtained using also the last output values (y_{k-j}). For MISO systems, the former equation is usually provided in the following compact notation

$$\hat{y}_k^p = (1 - A(q)) y_k + \sum_{m=1}^M B_m(q) u_{m,k-d_m} \quad (5)$$

where M is the number of inputs, $u_{m,k}$ is the m -th input discrete signal at time k , $A(q)$, $B_m(q)$ are polynomial in terms of the time-shift operator q^{-1} and d_m is the delay for the m -th input.

To estimate polynomial models, input delays and model orders have to be defined. If the physics of the system is known, the number of poles and zeros can be set accordingly. However, in most cases, the model orders are not known in advance. In those cases, the identification of an ARX model is a three steps procedure [21]: first the data have to be conditioned, then the structure of the model (i.e. polynomial's degrees and input delays) has to be identified, and finally the parameters (i.e. the polynomial's coefficients) have to be estimated.

In control design, the models are commonly used to simulate the behavior of the process, so that the previous outputs are not available. Hence, the final predictions \hat{y}_k will be obtained using the ARX model as

$$\hat{y}_k = \sum_{m=1}^M \frac{B_m(q)}{A(q)} u_{m,k-d} \quad (6)$$

in order to be compared with the LSTM ones (eq. 4).

2.3.1. Data conditioning

In order to identify a linear approximation around the working point of the system [22], a common precondition techniques involves the use of the deviation with respect to the working point instead of the regular signal. The signal average values are used to approximate the plant working point. Hence, the signals in hand become:

$$\begin{aligned} \delta y_k &= y_k - \bar{y} \\ \delta u_{m,k} &= u_{m,k} - \bar{u}_m \end{aligned}$$

where \bar{y} and \bar{u}_m are the average value of the output and of the m -th input signal. Then, the ARX model predictor became:

$$\hat{y}_k^p = \bar{y} + (1 - A(q)) \delta y_k + \sum_{m=1}^M B_m(q) \delta u_{m,k-d_m}$$

2.3.2. Model structure estimation

To get initial model orders and delays for the system, several ARX models can be estimated within a range of orders and delays, then the performance of these models can be compared. So, the model orders that correspond to the best model performance are selected and used as an initial guess for the next phase.

To guarantee a fair comparison to models having different degrees of freedom, a simple validation is used [23]. According to this techniques the data are split into two sets: identification and validation. Each model is identified using the first set while the performance of the model are evaluated on the second set. Then, the best model structure is the one minimising the prediction error computed on the validation data set.

The ASM model family [3], [4], [5] has been used to determine the ranges of the desired structure parameters. According to those models, the number of states ranges from 12 to 15, no delay is involved and the proposed model is strictly proper. Starting from this consideration, the parameter space was determined. The main idea is that the research space has to include the structure proposed by the physical models. In particular, the degree of $A(q)$ is connected to the order of the system then it was supposed $N_a > 15$, and for each input the sum delay and order should also include the same limit, $d_m + N_{bm} > 15$. Table 1 summarizes the proposed ranges.

Table 1. Compared ARX structures.

signal	component	range	selected
$O(t)$	order (N_a)	10 - 17	12
$q_{oss}(t)$	order (N_{b1})	2 - 12	6
	delay (d_1)	0 - 5	2
$q_{in}(t)$	order (N_{b2})	2 - 12	3
	delay (d_2)	0 - 10	5
$q_{ol}(t)$	order (N_{b3})	3 - 9	7
	delay (d_3)	0 - 10	6

2.3.3. Parameter estimation

This final phase is devoted to determine the model. The structure is now known, so the whole dataset can now be used to determine the best model. Moreover, this work aim to identify a model of a real plant which is intrinsically stable [3]. Then, the stability is enforced. From a mathematical point of view, this can be view as:

$$\arg \min_{A(q) \in H(N_a), B(q)} \sum_k \|y_k - \hat{y}_k^p\| \quad (8)$$

where $H(N_a)$ is set of the N_a -degree Hurwitz polynomial.

2.4. Datasets description

The WWTP is monitored through a Supervisory Control And Data Acquisition (SCADA) system used to manage the control protocols, alarm handling and also for data collection. From the SCADA interface the user can have access to sensors equipped in the chemical-physical reactor (where the wastewater is pre-treated), in the biological reactor and in the two membrane bioreactors, called respectively MBR1 and MBR2.

In this way, the signals of interest can be downloaded. The value of $q_{in}(t)$ is obtained from the chemical-physical reactor, from which it goes in input into the biological reactor, after the preprocessing of the wastewater. In the reactor are then present sensors that measure the ingoing oxygen flow rate, $q_{oss}(t)$, the oxygen concentration, $O(t)$, measured at a height of 4 meters. Then, the measure of the output liquid substrate flow rate, $q_{ol}(t)$, is obtained from MBR1 and MBR2, as the sum of the two components that from the biological reactor flow into the two membrane bioreactors.

Lastly, COD_{in} is obtained through laboratory analysis as well as $q_{om}(t)$ that is acquired daily by the operators who remove it from the reactor. These quantities are not considered in the models, not only for their constant values, but also for the difficulty in their acquisition.

Six different datasets are collected, each one containing six days of data, as in Table 2. All the datasets contain the three inputs and the output previously described. *Dataset 1* is used as testing dataset, while the remaining 5 datasets are used to train and validate the models.

Table 2. Datasets description.

Dataset name	Start date	End date
Dataset 1	06/12/2021	11/12/2021
Dataset 2	15/12/2021	20/12/2021
Dataset 3	21/12/2021	26/12/2021
Dataset 4	27/12/2021	01/01/2022
Dataset 5	01/03/2022	06/03/2022
Dataset 6	01/05/2022	06/05/2022

2.5. Performance Indexes

Three performance indexes are considered:

- Index of fitting (*FIT*): normalized index that indicates how much the prediction matches the real data. For a perfect prediction it is equal to 100% and it can also be negative. It is expressed as:

$$FIT = 100 \left(1 - \frac{\sum_{k=1}^S \|\hat{y}_k - y_k\|}{\sum_{k=1}^S \|y_k - \bar{y}\|} \right) \quad (9)$$

- Pearson correlation coefficient (ρ): it measures the linear correlation between two variables and has a value between -1 (total negative correlation) and +1 (total positive correlation).

$$\rho = \frac{\sum_{k=1}^S (y_k - \bar{y})(\hat{y}_k - \bar{\hat{y}})}{\|y_k - \bar{y}\| \|\hat{y}_k - \bar{\hat{y}}\|} \quad (10)$$

- Root Mean Squared Error (*RMSE*): it shows the Euclidean distance of the predictions from the real data using.

$$RMSE = \sqrt{\frac{1}{S} \sum_{k=1}^S (y_k - \hat{y}_k)^2} \quad (11)$$

where y is the vector of the real data, \hat{y} of the predicted ones, containing S elements each, and \bar{y} , $\bar{\hat{y}}$ are their respective mean values.

The training and validation of the models are performed on five datasets (Dataset 2 - Dataset 6) for a total amount of thirty days, while it is tested on the remaining one, Dataset 1.

3. Results

Due to privacy compliance, the working point of the plant is not made fully available: only the oxygen concentration is showed and available as full data ($\bar{O} = 1.3167 \text{ mg/l}$). This choice preserves the privacy of the company that provided them, and allows a complete comparison of the results.

The performance of the two approaches will be presented individually, and compared in the next Sections.

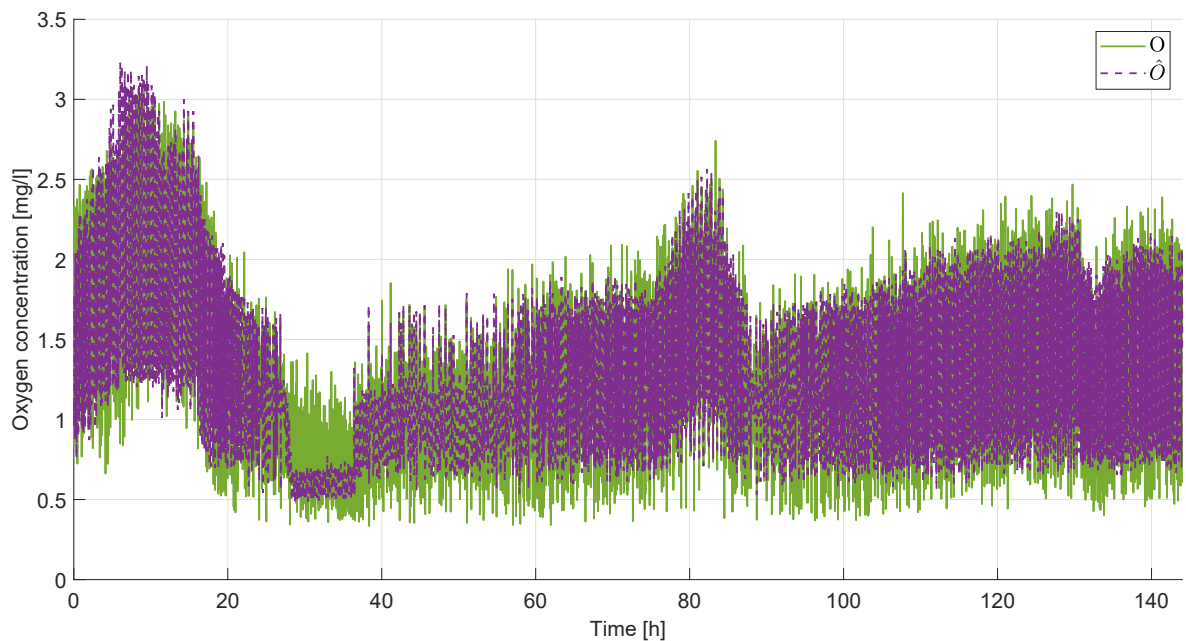
3.1. LSTM results

The LSTM network presents an architecture composed by a single layer LSTM with 160 neurons, trained for 1000 epochs, using Adam optimizer with learning rate $\alpha = 0.002$ and Mean Squared Error (MSE) loss function defined as:

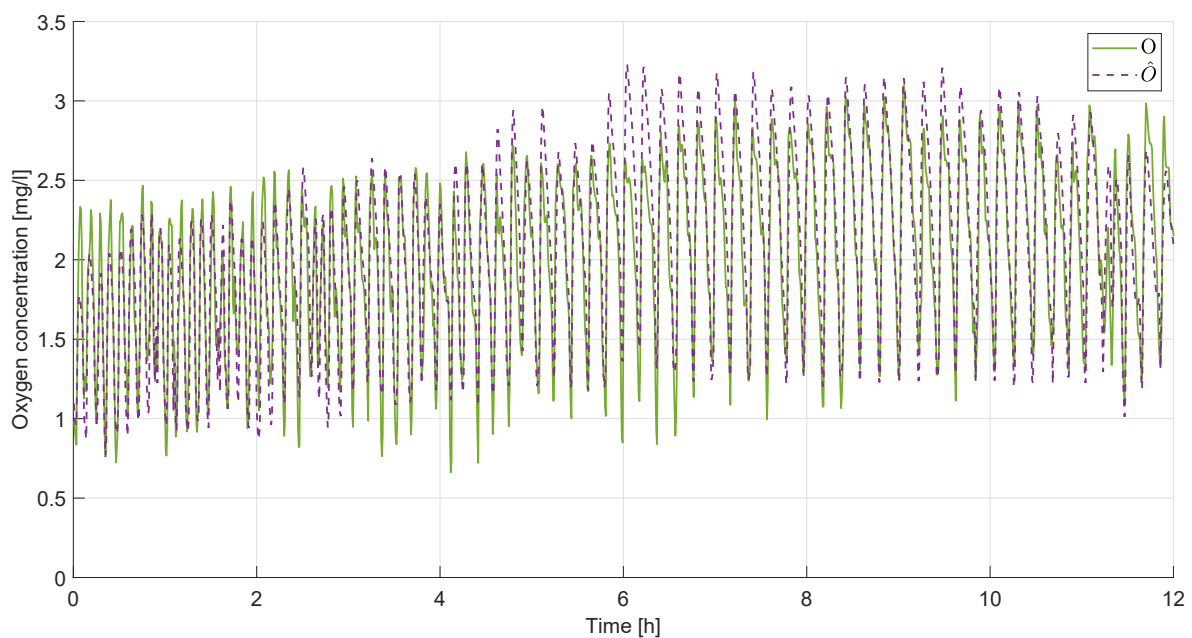
$$MSE = \frac{1}{S} \sum_{k=1}^S (\hat{y}_k - y_k)^2 \quad (12)$$

The values of the hyperparameters have been obtained through an optimization procedure with KerasTuner [24], employing the RandomSearch algorithm with the number of neurons spanning in the range [32; 512] and the learning rate in [0.001; 0.01].

The training of the network is performed on a computer Intel i9-10920X CPU with 3.50 GHz, with graphics processing unit GPU NVIDIA GeForce RTX 2080Ti; it has been written in Python 3.9, using TensorFlow [25], Keras API [26] and KerasTuner [24]. The results are quite satisfying, with $FIT = 60.56\%$, $\rho = 0.921$ and $RMSE = 0.206$. An example of the prediction obtained with the LSTM is reported in Figure 3, with a global overview in Figure 3a and a closer detail of the first 12 hours in Figure 3b, to better emphasize the accuracy of the result. It can be observed that both the global dynamic of the signal and the oscillations are correctly reproduced, with an overall satisfying outcome.



(a) Oxygen concentration



(b) First 12 h zoom

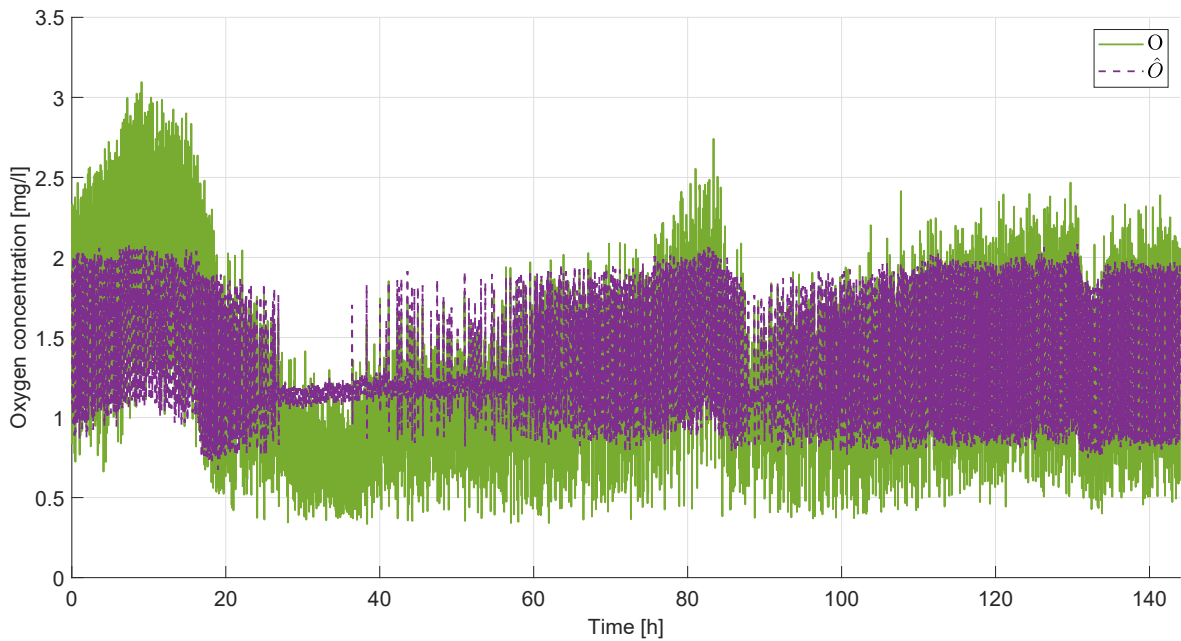
Figure 3. Results of the LSTM model.

3.2. ARX

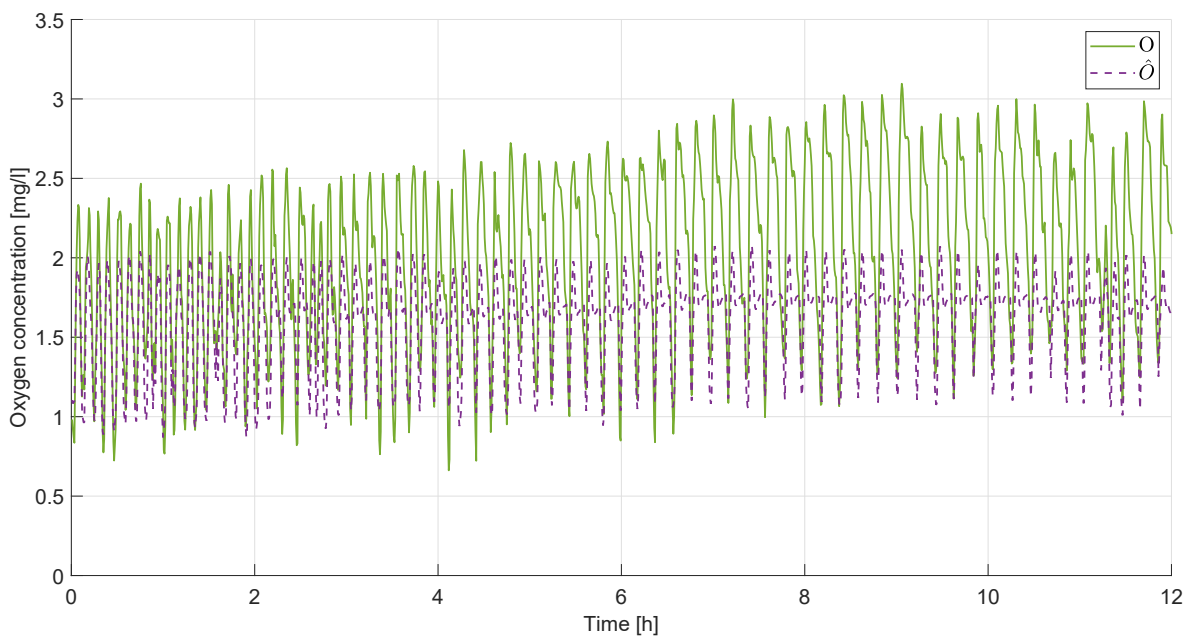
The procedure described in Section 2.3 was conducted using the System Identification Toolbox of Matlab [27].

Table 1 (last column) reports the structures of the model selected by the procedure.

The identified model was used to simulate the testing dataset. Figure 4a shows the comparison between the actual oxygen concentration and the predicted one. The oscillation frequencies and the overall behaviour are correctly estimated ($\rho = 0.833 > 0.7$), but the model fails in reaching the correct amplitude ($FIT = 41.20\% < 60\%$ and $RSME = 0.307$). Those results are emphasized in the zoom reported in Figure 4b and summarized by the performance indicators.



(a) Oxygen concentration



(b) First 12 h zoom

Figure 4. Results of the ARX model.

3.3. Discussion

The performance indexes are reported in Table 3 for both the presented techniques.

In Figure 5 a zoomed graph of the comparison of the two techniques is presented. It is evident how the LSTM is able to better follow the peaks while the ARX model underestimates in particular the maximum values.

Table 3. Model Comparison, results from Dataset1 simulation.

Index	ARX	LSTM
<i>FIT</i>	41.20%	60.56%
ρ	0.833	0.921
<i>RMSE</i>	0.307	0.206

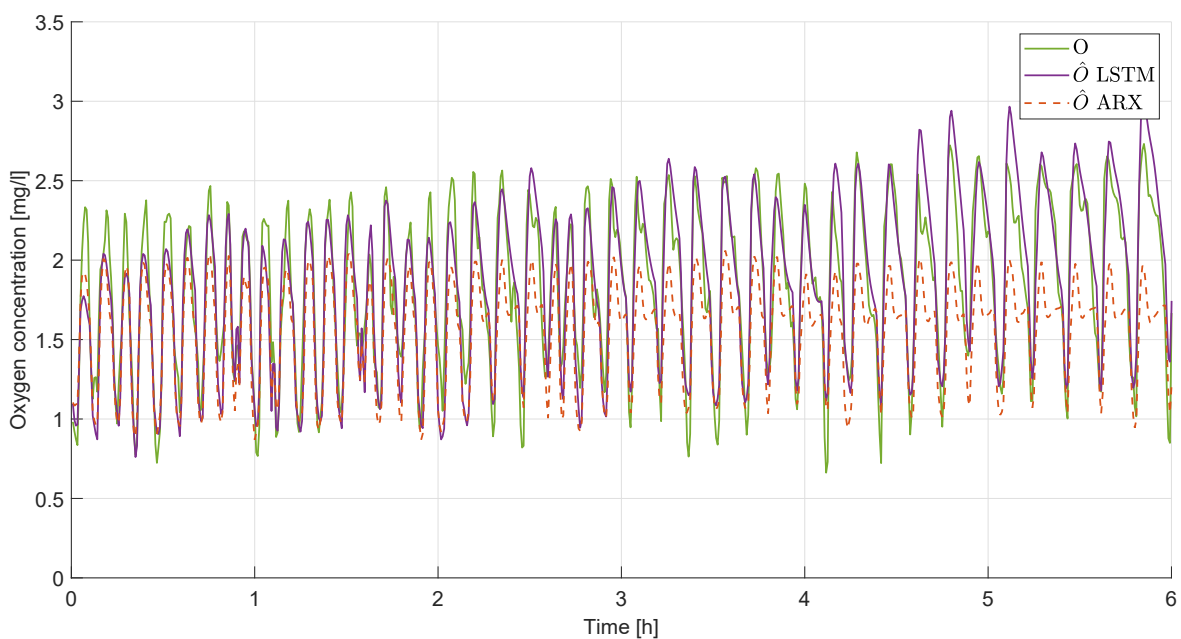


Figure 5. Comparison of the results obtained with the LSTM and the ARX model during the first 6 h.

4. Conclusions

This work proposes the use of a LSTM network to identify an input-output model suitable for the design of an oxygen concentration controller in an Activated Sludge process to treat municipal and industrial waste water. The model is identified using easy-accessible measures collected from a real plant located in Mortara, Italy. The LSTM showed satisfactory performances being able to catch both the oscillation and the overall behaviour ($\rho = 0.921$ and $FIT = 60.56\%$). It also overcomes the performance of a classic ARX ($\rho = 0.833$ and $FIT = 41.20\%$) which is not able to reach the correct amplitude.

The design of a controller based on the LSTM presented in this work is proposed as future development of this work.

Author Contributions: Conceptualization, C.T. and L.M.; methodology, C.T., F.D.P., F.I. and L.M. ; software, F.D.P. and F.I.; validation, C.T., F.D.P. and F.I.; formal analysis, C.T. and F.D.P.; resources, L.M.; data curation, F.D.P. and F.I.; writing—original draft preparation, C.T., F.D.P. and F.I.; writing—review and editing, C.T., F.D.P. and L.M.; supervision, C.T. and L.M.; project administration, L.M.; funding acquisition, L.M.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Azienda Servizi Mortara S.p.A..

Data Availability Statement: Data available in a publicly accessible repository at <https://arxiv.org/XXXX>.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASM1	Activated Sludge Model No. 1
ASM2	Activated Sludge Model No. 2
ASM3	Activated Sludge Model No. 3
ARX	AutoRegressive eXogenous
CAS	Conventional Activated Sludge
COD	Chemical Oxygen Demand
DEN	Denitrification
LSTM	Long Short-Term Memory
MSE	Mean Squared Error
NN	Neural Network
OX-NIT	Oxidation and nitrification
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
SCADA	Supervisory Control And Data Acquisition
TAMR	Thermophilic Aerobic Membrane Reactor
WWTP	WasteWater Treatment Plant

References

1. Hamitlon, R.; Braun, B.; Dare, R.; Koopman, B.; Svoronos, S.A. Control issues and challenges in wastewater treatment plants. *IEEE control systems magazine* **2006**, *26*, 63–69.
2. Gernaey, K.V.; Van Loosdrecht, M.C.; Henze, M.; Lind, M.; Jørgensen, S.B. Activated sludge wastewater treatment plant modelling and simulation: state of the art. *Environmental modelling & software* **2004**, *19*, 763–783.
3. Henze, M.; Grady Jr, C.L.; Gujer, W.; Marais, G.; Matsuo, T. A general model for single-sludge wastewater treatment systems. *Water Research* **1987**, *21*, 505–515. [https://doi.org/https://doi.org/10.1016/0043-1354\(87\)90058-3](https://doi.org/https://doi.org/10.1016/0043-1354(87)90058-3).
4. Gujer, W.; Henze, M.; Mino, T.; Matsuo, T.; Wentzel, M.; Marais, G. The Activated Sludge Model No. 2: Biological phosphorus removal. *Water Science and Technology* **1995**, *31*, 1–11. Modelling and Control of Activated Sludge Processes, [https://doi.org/https://doi.org/10.1016/0273-1223\(95\)00175-M](https://doi.org/https://doi.org/10.1016/0273-1223(95)00175-M).
5. Gujer, W.; Henze, M.; Mino, T.; Van Loosdrecht, M. Activated sludge model No. 3. *Water Science and Technology* **1999**, *39*, 183–193. Modelling and microbiology of activated sludge processes, [https://doi.org/https://doi.org/10.1016/S0273-1223\(98\)00785-9](https://doi.org/https://doi.org/10.1016/S0273-1223(98)00785-9).
6. Nelson, M.; Sidhu, H.S. Analysis of the activated sludge model (number 1). *Applied Mathematics Letters* **2009**, *22*, 629–635.
7. Nelson, M.; Sidhu, H.S.; Watt, S.; Hai, F.I. Performance analysis of the activated sludge model (number 1). *Food and Bioproducts Processing* **2019**, *116*, 41–53.
8. Gujer, W. Activated sludge modelling: past, present and future. *Water Science and Technology* **2006**, *53*, 111–119.
9. Sin, G.; Al, R. Activated sludge models at the crossroad of artificial intelligence—A perspective on advancing process modeling. *Npj Clean Water* **2021**, *4*, 1–7.
10. Hreiz, R.; Latifi, M.; Roche, N. Optimal design and operation of activated sludge processes: State-of-the-art. *Chemical Engineering Journal* **2015**, *281*, 900–920.
11. Collivignarelli, M.C.; Abbà, A.; Bertanza, G.; Setti, M.; Barbieri, G.; Frattarola, A. Integrating novel (thermophilic aerobic membrane reactor-TAMR) and conventional (conventional activated sludge-CAS) biological processes for the treatment of high strength aqueous wastes. *Bioresource technology* **2018**, *255*, 213–219.

12. Collivignarelli, M.; Abbà, A.; Frattarola, A.; Manenti, S.; Todeschini, S.; Bertanza, G.; Pedrazzani, R. Treatment of aqueous wastes by means of Thermophilic Aerobic Membrane Reactor (TAMR) and nanofiltration (NF): Process auditing of a full-scale plant. *Environmental monitoring and assessment* **2019**, *191*, 1–17.
13. Collivignarelli, M.C.; Abbà, A.; Bertanza, G. Why use a thermophilic aerobic membrane reactor for the treatment of industrial wastewater/liquid waste? *Environmental Technology* **2015**, *36*, 2115–2124.
14. Fraçz, P. Nonlinear modeling of activated sludge process using the Hammerstein-Wiener structure. In Proceedings of the E3S Web of Conferences. EDP Sciences, 2016, Vol. 10, p. 00119.
15. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
16. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: Continual prediction with LSTM. *Neural computation* **2000**, *12*, 2451–2471.
17. Gers, F.A.; Schraudolph, N.N.; Schmidhuber, J. Learning precise timing with LSTM recurrent networks. *Journal of machine learning research* **2002**, *3*, 115–143.
18. Pascanu, R.; Mikolov, T.; Bengio, Y. On the difficulty of training recurrent neural networks. In Proceedings of the International conference on machine learning. PMLR, 2013, pp. 1310–1318.
19. Greff, K.; Srivastava, R.K.; Koutník, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* **2016**, *28*, 2222–2232.
20. Pham, V.; Bluche, T.; Kermorvant, C.; Louradour, J. Dropout improves recurrent neural networks for handwriting recognition. In Proceedings of the 2014 14th international conference on frontiers in handwriting recognition. IEEE, 2014, pp. 285–290.
21. Ljung, L., Ed. *System Identification (2Nd Ed.): Theory for the User*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1999.
22. Söderström, T.; Stoica, P. *System Identification*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1989.
23. Godfrey, K. Correlation methods. *Automatica* **1980**, *16*, 527–534. [https://doi.org/https://doi.org/10.1016/0005-1098\(80\)90076-X](https://doi.org/https://doi.org/10.1016/0005-1098(80)90076-X).
24. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; et al. KerasTuner. <https://github.com/keras-team/keras-tuner>, 2019.
25. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org.
26. Chollet, F.; et al. Keras. <https://keras.io>, 2015.
27. The MathWorks Inc., Natick, Massachusetts, United States. *System Identification Toolbox version: 9.16 (R2022a)*, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.