

A Recurrent Neural Network for Wastewater Treatment Plant effluents' prediction

Ivan Pisa, Ignacio Santín, Jose Lopez Vicario, Antoni Morell, Ramon Vilanova,
 {Ivan.Pisa, Ignacio.Santin, Jose.Vicario, Antoni.Morell, Ramon.Vilanova }@uab.cat
 Department of Telecommunications and Systems Engineering,
 Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain

Abstract

Wastewater Treatment Plants (WWTP) are industries devoted to process water coming from cities' sewer systems and to reduce their contamination. High-pollutant products are generated in the pollutant reduction processes. For this reason, certain limits are established and violations of them are translated into high economic punishments and environmental problems. In this paper data driven methods are performed to monitor the WWTP behaviour. The aim is to predict its effluent concentrations in order to reduce possible violations and their derived costs. To do so, an alarm generation system based on the application of Artificial Neural Networks (ANNs) is proposed. The proposed system shows a good prediction accuracy (errors around 5%) and a reduced miss-detection probability (30%).

Keywords: Wastewater Treatment Plants, Neural Networks, LSTM cells, BSM2

1 Introduction

Wastewater Treatment Plants (WWTPs) are devoted to manage and process the residual waters coming from the cities' sewer systems. In WWTP plants, biological and biochemical processes are applied to water to decrease the pollutants such as ammonium (S_{NH}) and total nitrogen (S_{Ntot}) concentrations. The processes, which are highly complex, are also applied to avoid the contamination of the ecosystem where treated water flows. Furthermore, there exist some limits established by law which determine the maximum concentration levels of certain WWTP products. Violations of these limits are translated into high economic punishments and environmental problems.

In that sense, the International Water Association (IWA) has developed a mathematical model, the Benchmark Simulation Model No.1 (BSM1) [2], which replicates the WWTP's real behaviour. It models the biological treatment of the water based on the Activated Sludge Model No.1 (ASM1) [7]. BSM1 has been enhanced with the appearance of

the Benchmark Simulation Model No.2 (BSM2) which includes the activated sludge treatment [4]. Both, BSM2 and BSM1 models, are adopted as the evaluation scenarios of new control strategies to maintain the pollutants below the established limits.

In such a context, the appearance of Artificial Neural Networks (ANNs) [5] has arisen as a method able to model the WWTP behaviour without the knowledge of BSM2 and BSM1 models, they only require the WWTP's influent and effluent data [8].

For instance, ANNs are adopted to forecast or predict some parameters to feed a control strategy. In that sense, Foscoliano *et.al* propose the adoption of RNN networks to predict the WWTP's nutrient concentrations and then, feed a Model Predictive Control (MPC) strategy to assure that concentrations are under the established limits [3]. Furthermore, Santín *et.al* adopt two Multilayer Perceptron (MLP) neural networks to predict the ammonia and total nitrogen concentrations in the effluent and determine whenever a violation of their limits is likely to occur [10].

On the other hand, there are several works where ANNs have been adopted to monitor or predict some WWTPs' parameters. For instance, in [6], the Chemical Oxygen on Demand (COD), the Suspended Solids (SS) and the aeration tank concentrations (S_O) are predicted by means of three different Multilayer Perceptron (MLP) neural networks yielding a Mean Absolute Percentage Error (MAPE) in the prediction around 4.48%.

In this work we propose a system based on ANNs to predict violations of the ammonium and total nitrogen concentrations and generate an alarm to determine when the control strategy (MPC + Fuzzy Logic) has to be activated. Furthermore, the temporal behaviour and dependence between WWTP's inputs and outputs will be modeled adopting LSTM cells. Our main contributions to the state of the art are:

- Adopt RNN and LSTM structures to deal with time series and temporal data.

- Consider online and continuous data. Thus, the time dependence is conserved and predictions can be performed with information gathered in the same WWTP.
- Reduce the time between predictions. As a consequence of the adoption of continuous data, predictions are given every 15 minutes.

2 Problem Definition

The main objectives of WWTP plants is to reduce the amount of pollutant concentrations in water and return it to its natural cycle. Consequently, one of the main objectives of this work consists in the adoption of data driven methods to model the WWTPs behaviour in order to predict effluent nutrient concentrations. With these predictions, possible violations of the established effluent limits for such nutrients can be eliminated, therefore reducing the receiving water's pollution and avoiding possible economic sanctions. The considered pollutants are the nitrogen and phosphorous.

In this work we will concentrate on the nitrogen removal. Nitrogen is considered in the form of ammonia and total nitrogen, treated with Biological WWTPs. The Biological treatment is split into two processes which are known as nitrification and denitrification. The first one is in charge of transforming the ammonia into nitrate while the second converts the nitrate into nitrogen and other gaseous products [14]. These processes are modeled by the widely accepted ASM models [7], that are incorporated into the BSM2, benchmark scenario developed by the IWA [4].

Table 1: Effluent quality limits

Effluent quality limits	
Variable	Value
S_{Ntot}	<18 mg/L
S_{NH}	<4 mg/L

For what matters to the BSM2 scenario, the established effluent limits are shown in Tab.1 where S_{Ntot} and S_{NH} are the total nitrogen and the ammonium concentrations [4], both measured in the effluent.

3 Material & Methods

3.1 Benchmark Simulation Model No. 2

The adopted input and output data will be generated through the usage of the BSM2 model, which is an enhancement of BSM1 model [2]. BSM2 is formed by two clearly differentiated sections, the

first one is in charge of the water's biological treatment at the activated sludge reactors, and the second one is in charge of the sludge treatment. Water's biological treatment is based on BSM1.

BSM1 consists in five biological reactor tanks which are connected in series. The first two tanks are anoxic tanks (they have a lack of oxygen) whereas the three remaining are aerobic. Their biological process is described in [7]. An internal recycle flow is considered from the last tank to the first one. In addition, BSM1 model considers a clarifier connected to the fifth tank's output. This clarifier is adopted to perform the sludge sedimentation process following the Takács model [12]. Concerning the flow rates, the plant considered in BSM2 and BSM1 models is designed considering an average flow rate of $20648.36 \text{ m}^3/d$. Each anoxic tank has a volume of 1500 m^3 and each aerobic tank has a volume of 300 m^3 . Thus, the retention time considering the total volume of the tanks and the flow rate is about 14h.

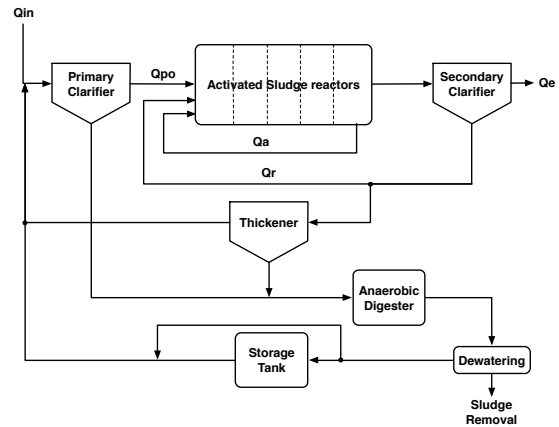


Figure 1: BSM2 model for a plant structure

BSM2 adds different elements whose purpose are to treat the sludge. In such a context, a primary clarifier, a thickener for the sludge wasted in the secondary clarifier, an anaerobic digester to treat the wasted solids and a dewatering module are added to the BSM1 model as it is shown in Fig.1. In it, the different flow rates are specified: Q_{in} corresponds to the input flow, Q_{po} is the primary clarifier overflow, Q_a is the internal recycle flow rate, Q_r is the sludge internal recycle flow rate and finally Q_e is the effluent flow rate.

The influent of the BSM2 simulation model is defined as an unique input regarding 609 days. This influent takes into account dry, rainy and stormy weather. In addition, temperature variations are registered. However, before the input influent the simulator has to be calibrated in close-loop by means of a 200 days input file of constant inputs. After the calibration process, the 609 days

influent is adopted. Nonetheless, the BSM2 simulation protocol specifies that only results from days 245 to 609 have to be taken into account for plant evaluation. The BSM2 effluent file is generated adopting the plant structure and the control strategies adopted. It gathers information about the nutrient concentrations and the effluent rate among others with a sampling time of 15 minutes. Thus, four consecutive samples correspond to one hour of WWTP's operation.

3.2 Artificial Neural Networks

The ability of ANNs to model non-linear system has motivated us to adopt them as the main method to generate WWTPs' effluent predictions. ANNs structures consist in a set of layers [5]. Each layer presents an amount of hidden neurons which are characterized by the activation function they adopt. In this context, activation functions are applied to obtain the neuron output in terms of the input. Layers are connected to each other through weights and biases.

ANN's model is generated by means of an iterative process called training process where the values of the weights and biases are optimized. In that sense, weights and biases are modified iteratively taking into account the input data and the generated outcomes. For this reason, a large dataset of input and output values is required. Moreover, this dataset will be split in three subsets following the next otherwise usual distribution: 70%, the training dataset devoted to perform the training process; 15%, the validation dataset to validate the training and to see if overfitting is committed; and 15%, the test dataset to see the performance of the network with new data.

The values of weights and biases are optimized taking into account a cost function. It corresponds to a measurement of how well is predicting or classifying the ANN. For this reason, it is selected according to the purpose of the neural network. Since this work is focused on regression purposes, the Mean Square Error (MSE) function (Eq. 1) is adopted,

$$J(\theta) = \frac{1}{N} \sum_i^N (y_i - \hat{y}_i)^2 \quad (1)$$

where N is the number of examples, y_i is the desired output, \hat{y}_i corresponds to the ANN's prediction and θ corresponds to the weights of the neural network.

In this work, predictions are performed adopting LSTM cells which are a type of Recurrent Neural Networks (RNNs). RNNs are ANNs wherein

connections between hidden neurons form a directed cycle (unfolding process is observable in Fig.2). This means that the hidden state of the neural network is adopted to give some extra information about the dynamics of the inputs. In addition, RNNs are widely known due to its behaviour modeling sequences and continuous signals such as text or voice [13]. In addition, they are also recommended due to their ability modeling systems with non-linear dynamics [5, Chapter 12]. Furthermore, RNN structures have yielded very good results dealing with time series and sequence recognition problems [11]. This motivates us to consider them in our problem, the signals we are dealing with are continuous in time with a strongly dependence between inputs and outputs.

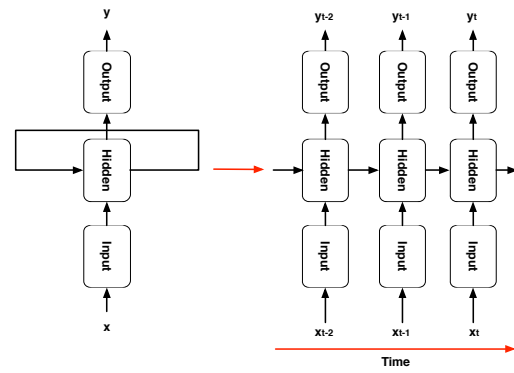


Figure 2: Unfolding process for RNN Network for Time Series

Training process is performed adopting the back-propagation through time (BPTT) algorithm, which is based on the iterative update of the network weights towards the opposite of the gradient of the loss function w.r.t the weights. However, back propagation algorithm has some drawbacks when it is adopted in the training process of RNNs, due to the well-known exploding or vanishing gradients problem [1]. In that sense, gated networks such as the LSTM cell approach (See Fig.3) appeared as a promising alternate structure to alleviate this problem. They replace the hidden neuron by a memory cell with three gates, the forget, the input and the output gates. The forget gate can reset the memory cell which will be fulfilled with the information coming from the input gate. The output gate will determine if the output will influence on other cells [5, Chapter 10].

In order to avoid overfitting, some regularization techniques have been applied to LSTM cells. Among the different techniques, the three adopted are Early Stopping, L2 and DropOut regularization techniques [5, Chapter 7]. Early Stopping consists in the stop of the training process execution when performance of the generated model

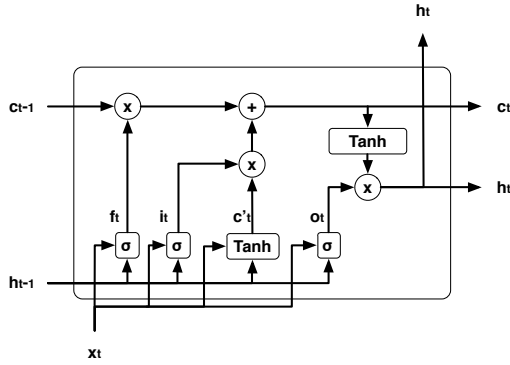


Figure 3: LSTM Cell

starts to drop and become worse than before. L2 consists in adding extra penalty to the weights optimization. As a consequence, it is harder for the network to match the training examples as for weights cannot grow without penalty [5, Chapter 7]. Dropout consists in the application of “keep probabilities” at the output of the hidden neurons in a layer, i.e, neurons’ output are activated following the given “keep probabilities”. It can be interpreted as performing a mixture of models as far as different networks are considered (depending on the set of active neurons). This leads to a generalization of the weights.

4 Effluent nutrient violation’s prediction in WWTP’s

As it has been stated in previously sections, the main purpose of WWTP plants is to decrease the pollutant levels and therefore avoid the contamination of the receiver waters. Thus, the effluent’s levels are limited and therefore any violation of them incur into an economical sanction to the plant. In addition, there are several control strategies proposed in the literature whose objectives are to reduce the levels of effluent products. However, the application of control strategies can involve an increase of the plant’s operational cost.

In such a context, we propose the ANN-based Effluent Violation Prediction (AEVP) system (See Fig.4) which objective is to predict possible violations of WWTP’s effluent limits, thereby the control strategies can be applied only when they are required and also the effluent limit violations can be avoided. The system consists in three different stages: the Gathering Information, the ANN (based on RNNs) and the Alarm Generation subsystems. Their purpose is to gather the online information from the WWTP plant, preprocess it and then perform predictions of the effluent levels by means of ANNs. Predictions will be contrasted

with the limits (See Tab.1).

4.1 Prediction Objective & Input Data

ANNs will be adopted in the AEVP system to model the WWTP’s dynamics and behaviours in an attempt to overcome the use of the high complexity deterministic models considered in BSM2 [4]. Generated models will be used to predict the effluent concentrations of ammonium and total nitrogen ($S_{NH,e}$ and $S_{Ntot,e}$). These predictions will determine when the effluent concentration limits (See Table 1) will be violated or, more precisely, they are at risk of being violated. The ANN will be trained taking into account a subset of the WWTP influent data and its correspondent outputs.

In this work, the inputs gathered by the Gathering Information subsystem consist in online data given by sensors placed in the WWTP plant. In that sense, predictions can be performed directly without the necessity of laboratory analysis as it happens with offline data [15]. The influent subset corresponds to:

- $S_{NH,po}$: Ammonium concentration of the primary clarifier
- Q_{po} : Overflow rate of the primary clarifier
- $S_{NH,po} \cdot Q_{po}$
- Q_{rin} : internal recycle flow rate
- T_{as} : Temperature

$S_{NH,po}$ is considered because it is one of the pollutants with the higher influence on the total final nitrogen concentration ($S_{Ntot,e}$). It not only increases the concentration of ammonium in the effluent, but it also affects the nitrification process and consequently the resulting $S_{Ntot,e}$. T_{as} is added due to its influence in the nitrification and denitrification processes performed in the tanks of the BSM2 model. The product between $S_{NH,po}$ and Q_{po} , is also considered due to its appearance in the general mass balancing equations of the ASM1 model [7]. The output data has been gathered from BSM2 models with the MPC + Fuzzy Logic strategy defined in [9]. Data has been normalized to zero mean and unit variance to reduce the range of input values.

In order to organize the different time measures, a sliding window has been implemented in the Gathering Information subsystem to gather the input data and pass it through the Neural Network structures. In that sense, the sliding window has a window length (WL), defined as the record data history considered for prediction, and a prediction

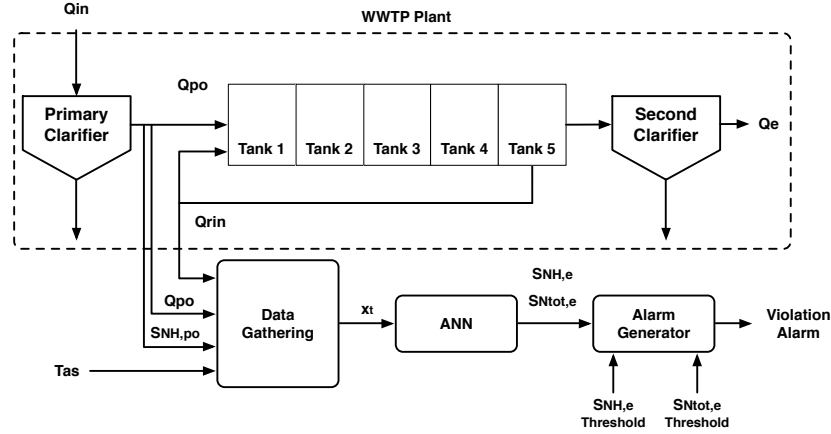


Figure 4: AEVP system to predict WWTP's effluent levels and determine if violations will be committed

horizon (PH), where PH is defined as the amount of time the prediction can be given in advance. Their configurations are the following ones:

- Window Length (WL): A WL of 10 hours (40 samples) is considered. The input data not only corresponds to the values seen at each sampling time, but also to those seen previously. Thus, when the sliding window slides the oldest input data is discarded and the newest one is considered.
- Prediction Horizon (PH): A PH of 4 hours (16 samples) is considered. Output data corresponds to the effluent concentrations of total nitrogen ($S_{N_{tot,e}}$) and ammonium ($S_{NH,e}$) which will be observable in four hours.

The adoption of these WL and PH configurations are motivated by the fact that the retention time of the BSM2 model is estimated of 14 hours. Thus, the first input value of the sliding window is supposed to generate the output prediction after 14 hours. For example, if it is gathered at 00:00h, the generated output is supposed to come out of the WWTP at 14:00h. Thus, if a sliding window of 10h is considered, a PH of 4h is needed (See Fig.5).

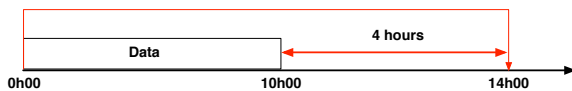


Figure 5: Sliding window example

4.2 ANN's prediction structures

Predictions are performed according to the ANN subsystem which is in charge of taking the input values given by the Gathering Information subsystem and predicting the effluent concentration levels. Those correspond to the ammonium ($S_{NH,e}$)

and total nitrogen ($S_{N_{tot,e}}$) concentrations. In that sense, LSTM Structures are adopted as a prediction approach: five LSTM-based prediction structures, three for the effluent ammonium concentration ($S_{NH,e}$) and two for the total nitrogen effluent concentration ($S_{N_{tot,e}}$). The proposed LSTM-based prediction structures are:

- Ammonium Prediction Structure (APS): The APS structure is formed by an input layer of 5 input neurons connected to the LSTM hidden layers (one (APS-LSTM1), two (APS-LSTM2), or three (APS-LSTM3)) of 50 hidden units each LSTM hidden layer. Each hidden layer contains one LSTM cell. The last hidden layer is connected to an output layer formed by one output neuron with a linear activation function.
- Total Nitrogen Prediction Structure (TNPS): The TNPS structure is formed by an input layer of 5 input neurons connected to the LSTM hidden layers (one (TNPS-LSTM1) or two (TNPS-LSTM2)) of 10 hidden units each LSTM hidden layer. The last LSTM hidden layer is connected to an output layer with one output neuron adopting the linear activation function. TNPS-LSTM2 is presented in Fig.6.

4.3 Prediction & Alarm Generation Results

System's performance is evaluated by means of three different criteria: the Mean Absolute Percentage Error (MAPE), the false alarm (P_{fa}) and the miss-detection (P_{miss}) probabilities. The Alarm Generation contrasts the prediction with the system's threshold, therefore if predictions are higher than alarm thresholds (violation is prone to be committed) an alarm is generated. Thus, P_{fa}

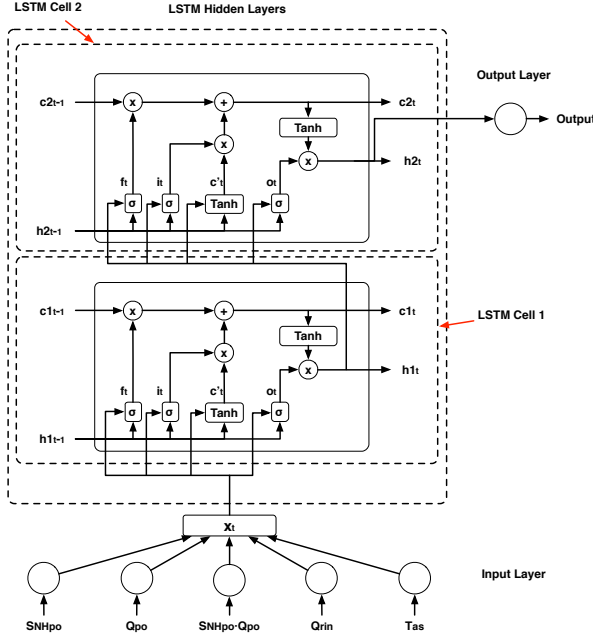


Figure 6: Total Nitrogen Prediction Structure (TNPS-LSTM2) - 2 LSTM hidden layers with 10 hidden units per LSTM cell

is defined as the probability of predicting an inexistent limit violation whereas P_{miss} is defined as the probability of missing a real limit violation.

The MAPE criteria is computed as follows:

$$MAPE = \frac{1}{N} \cdot \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100 \quad (2)$$

where MAPE represents the percentage of error committed in the predictions with respect to the real output value. The main point is not to predict the exact behaviour of the WWTP plant with high precision but to detect the possible effluent violations.

Finally, predictions are adopted to feed the Alarm Generation subsystem. It will compare the predicted value with the effluent limits. Thus, probabilities of false alarm (P_{fa}) and miss-detection (P_{miss}) make sense in the evaluation of the Alarm Generation Subsystem's results. P_{miss} and P_{fa} are defined as:

$$P_{miss} = P(\hat{y}_i < \gamma | y_i \geq \gamma) \quad (3)$$

$$P_{fa} = P(\hat{y}_i \geq \gamma | y_i < \gamma) \quad (4)$$

where \hat{y}_i corresponds to the i th value of the predictions, y_i represents the i th real output and γ consists in the effluent limit or threshold. The ef-

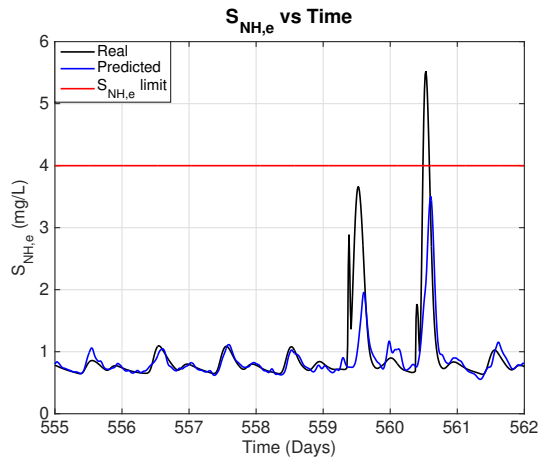
fluent limits adopted to determine if there exists an alarm are the ones shown in Tab.1.

Table 2: Performance of ANN effluent concentration prediction

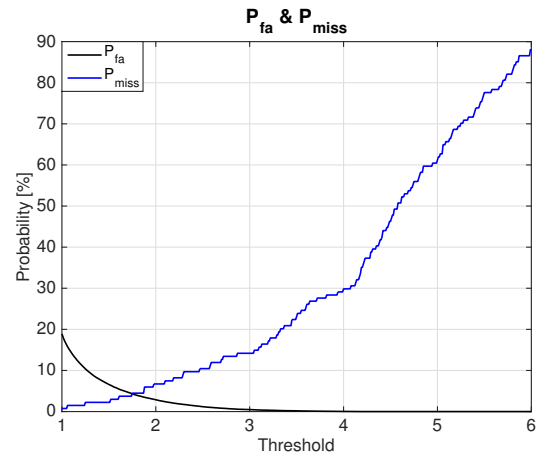
ANNs prediction performance			
<i>Ammonium in the effluent - $S_{NH,e}$</i>			
Structure	MAPE [%]	P_{fa} [%]	P_{miss} [%]
APS-LSTM1	8.08	0.01	41.04
APS-LSTM2	7.93	0.03	29.85
APS-LSTM3	7.87	0.04	35.07
<i>Total Nitrogen in the effluent - $S_{Ntot,e}$</i>			
Structure	MAPE [%]	P_{fa} [%]	P_{miss} [%]
TNPS-LSTM1	4.84	0.45	39.57
TNPS-LSTM2	4.69	0.44	31.75

Performance of the ammonium and total nitrogen prediction structures is shown in Tab.2. MAPE criteria shows that the percentage of the error committed in the prediction is lower than a 10% for $S_{NH,e}$ prediction structures. In such a context, one can observe that the lowest MAPE is given by the APS with 2 hidden layers structure showing a P_{miss} of 29.85%. However, there is an exception with the APS with 3 hidden layers. It shows a lower MAPE but a higher P_{miss} than the APS with 2 hidden layers. This effect is motivated by the fact that APS with 3 HL is affected by overfitting. Overfitting is defined as the effect where the ANN's model is not able to generalize. It is produced when the model is over trained. Concerning the $S_{Ntot,e}$ prediction structures, one can observe that MAPE criteria shows that the error committed is lower. In terms of P_{miss} , the same tendency as in $S_{NH,e}$ is observable. Finally, it can be stated that there exists a correlation between the MAPE and the P_{miss} criteria, when the former decreases the latter does so. Performed predictions are observable in Fig.7. Finally, APS and TNPS with 2HL structures (APS-LSTM2 & TNPS-LSTM2) are the ones showing the best performance in terms of MAPE & P_{miss} (See Tab.2).

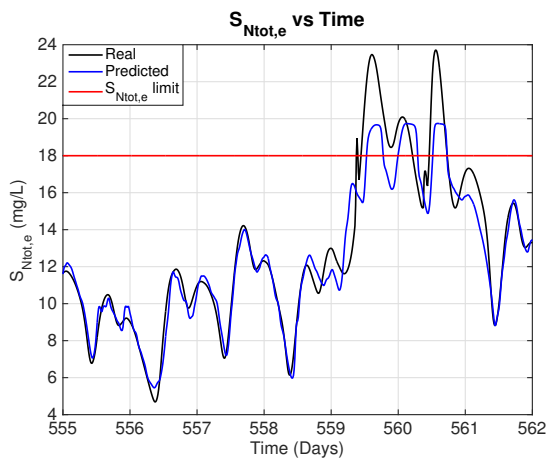
In terms of the predictions' accuracy, miss-detection can be reduced by adjusting violation threshold, which could lead to an increase of the false alarm probability. In other words, the $S_{NH,e}$ and $S_{Ntot,e}$ limits of the alarm generation process can be reduced but maintaining the real limits to those shown in Tab.1. Thus, the Receiver Operating Characteristic (ROC), P_{fa} and $P_{detection}$ (being $P_{detection} = 1 - P_{miss}$), can be obtained as a function of the threshold of the alarm generation system. By modifying the threshold, one can adjust P_{fa} and P_{miss} . For instance, ROC curve shows that a $\gamma = 3 \text{ mg/L}$ yields a $P_{fa} = 0.44\%$ and a $P_{detection} = 85.82\%$ whereas, if a $\gamma = 1 \text{ mg/L}$, $P_{detection} = 99.25\%$ and $P_{fa} = 18.9\%$. No-



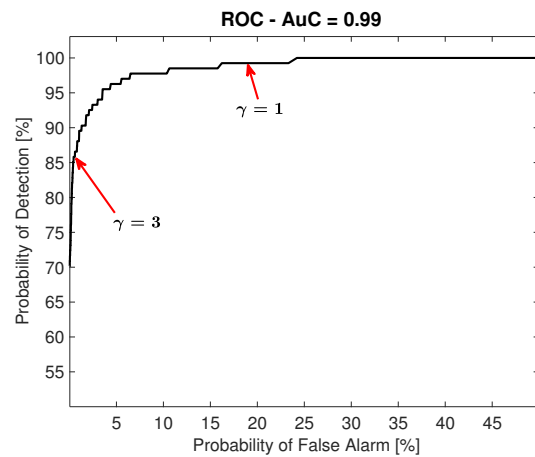
(a) NH concentration's prediction



(a) Probability of False alarm and Miss-detection



(b) Ntot concentration's prediction



(b) Receiver Operating Characteristic (ROC)

Figure 7: Real and predicted NH and Ntot concentrations in the effluent - APS-LSTM2 & TNPS-LSTM2 structures are adopted

tice that, curve of ROC showing an Area under the Curve (AuC) of 0.99 is obtained (a value equal to 1 represents perfect performance). P_{fa} vs. P_{miss} for APS-LSTM2 configuration is shown in Fig.8.

5 Conclusion

In this paper, the application of neural networks to predict $S_{NH,e}$ and $S_{Ntot,e}$ in WWTP's effluent is presented. Their main goal is to determine when effluent concentrations are prone to exceed the established limits. In that sense a system to generate alarms has been proposed. Its purpose is to predict future WWTP's concentration values adopting RNN in order to reduce the number of violations and the derived costs.

Predictions are performed adopting two RNN-based LSTM structures, one for $S_{NH,e}$ and another for $S_{Ntot,e}$. They will determine when effluent concentrations limits are at risk of being

Figure 8: P_{fa} vs. P_{miss} (a) and ROC (b) for APS-LSTM2. ROC shows the trade-off between P_{fa} and $P_{detection}$. $P_{detection} = 1 - P_{miss}$.

violated and consequently when an alarm has to be generated. Thereby control strategies will be executed when alarms are generated. This can be translated in a reduction of the operation cost of the plant and also the violations of effluents' limits.

Results show a miss-detection probability (P_{miss}) around 30% for both cases. In addition, low prediction errors (low MAPE) are achieved. Moreover, the system considers the option to modify alarm thresholds in order to adjust the miss-detection probability at the expense of a higher false alarm. ROC curve has been computed showing an Area under the Curve of 0.99 (1 is perfect performance).

Acknowledgement

This work has received the support from the Spanish Ministry of Economy and Com-

petitiveness program under MINECO/FEDER grant DPI2016-77271-R and also from La Secretaria d'Universitats i Recerca del Departament d'Empresa i Coneixement de la Generalitat de Catalunya i del Fons Social Europeu under FI grant. Authors belong to the Recognized Consolidated Research group SGR 1202 and to the Recognized Emergent Research group SGR 1670 by the Catalan Government.

References

- [1] Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, mar 1994.
- [2] J. B. Copp, *The cost simulation benchmark: Description and Simulator Manual (cost action 624 and action 682)*. Luxembourg: Office for Official Publications of the European Union, 2002.
- [3] C. Foscoliano, S. Del Vigo, M. Mulas, and S. Tronci, "Predictive control of an activated sludge process for long term operation," *Chemical Engineering Journal*, vol. 304, pp. 1031–1044, 2016.
- [4] K. V. Gernaey, U. Jeppsson, P. A. Vanrolleghem, and J. B. Copp, *Benchmarking of control strategies for wastewater treatment plants. Scientific and Technical Report No.23*. London, UK: IWA Publishing, 2014.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [6] D. Güçlü and Ş. Dursun, "Artificial neural network modelling of a large-scale wastewater treatment plant operation," *Bioprocess and Biosystems Engineering*, vol. 33, no. 9, pp. 1051–1058, 2010.
- [7] M. Henze, L. Grady Jr, W. Gujer, G. V. R. Marais, and T. Matsuo, "Activated sludge model no 1," *Scientific and Technical Report No.1*, vol. 29, 01 1987.
- [8] H. S. Ou, C. H. Wei, H. Z. Wu, C. H. Mo, and B. Y. He, "Sequential dynamic artificial neural network modeling of a full-scale coking wastewater treatment plant with fluidized bed reactors," *Environmental Science and Pollution Research*, vol. 22, no. 20, pp. 15 910–15 919, 2015.
- [9] I. Santín, C. Pedret, and R. Vilanova, "Fuzzy Control and Model Predictive Control Configurations for Effluent Violations Removal in Wastewater Treatment Plants," *Industrial and Engineering Chemistry Research*, vol. 54, no. 10, pp. 2763–2775, mar 2015.
- [10] I. Santín, C. Pedret, R. Vilanova, and M. Meneses, "Advanced decision control system for effluent violations removal in wastewater treatment plants," *Control Engineering Practice*, vol. 49, no. 2, pp. 60–75, 2016.
- [11] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, nov 2017.
- [12] I. Takács, G. G. Patry, and D. Nolasco, "A dynamic model of the clarification-thickening process," *Water Research*, vol. 25, no. 10, pp. 1263–1271, 1991.
- [13] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in *Lecture Notes in Computer Science (including sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10253 LNCS. Springer, Cham, jun 2017, pp. 477–492.
- [14] Z. Yang, PY and Zhang, "Nitrification and denitrification in the wastewater treatment system," in *Traditional technology for environmental conservation and sustainable development in the asian-pacific region*, 1995, pp. 145–158.
- [15] L. J. Zhang, N. Li, J. J. Zhang, and X. Y. Tian, "Application of neural network in modeling of activated sludge wastewater treatment process," in *2017 36th Chinese Control Conference (CCC)*. IEEE, jul 2017, pp. 4556–4561.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC-BY-NC 3.0 license (<http://creativecommons.org/licenses/by-nc/3.0/>).