Adaptive Neural Network Based Monitoring of Wastewater Treatment Plants

Mohammed S. Alharbi¹, Pei-Ying Hong² and Taous-Meriem Laleg Kirati¹

Abstract—The quality of the treated wastewater is conditioned by the performance of wastewater treatment processes. However, real-time monitoring of non-easily measured quality variables in wastewater treatment plant (WWTP) is a challenging problem. In this paper, an adaptive online monitoring approach that is based on a long short term memory (LSTM) neural networks is proposed to estimate the bacterial concentration, mixed liquor suspended solids (MLSS) and mixed liquor volatile suspended solids (MLVSS) in WWTP. Due to lack of a large dataset and difficulties to experimentally measure quality variables, a Wasserstein generative adversarial network with gradient penalty (WGAN-GP) is designed to generate synthetic data for training. A tuned hyperparameters are obtained for the proposed method. In addition, the performance is compared with the traditional LSTM using two datasets. Finally, the results indicate that WGAN successfully generates realistic training samples and an quality variables are monitored with satisfactory performance.

I. INTRODUCTION

The main goal of a wastewater treatment plant (WWTP) is to reduce pollutant concentrations as it disrupts the transmission of infectious pathogenic vectors as deadly viruses and bacteria. In addition, it allows to obtain clean treated wastewater that can be reused to alleviate water scarcity. In WWTP, the activated sludge process (ASP) is a commonly used biological treatment method where active microorganisms are responsible for the organic and nutrient removal from wastewater. Hence, the amount of microorganisms is a critical quantity for WWTP operation and design. Also, part of the bacteria in the untreated wastewater and settleable solids, settle down when the water is recycled in the activated sludge tank and in the clarifiers as an activated sludge blanket. Having a high bacterial concentration in the final product water is detrimental to the environment and public health. Therefore, WWTPs must be monitored for their effectiveness in purifying wastewater and ability to remove bacterial contaminants.

However, there are essential quality key variables in WWTP that are not easily measured such as the bacteria cell concentration, mixed liquor suspended solids (MLSS) and mixed liquor volatile suspended solids (MLVSS) that correspond to microorganisms concentration. The traditional ways to measure bacterial concentrations require sample taking and laboratory analysis like counting or using flow cytometry but these methods have a time-lag problem and cannot reflect the state of the process in real-time. For instance, active biomass concentration requires cultivation to enumerate, which is tedious and time-consuming, whereas the flow cytometry experience is essential to operate the instrument and determine the bacteria concentration [16]. Thus, an effective way to solve this problem and benefits the plant operators is the design of an estimation method that can accurately estimate the non-easily measurable variables in WWTP. Moreover, estimation methods have been widely developed for WWTP estimation and monitoring purposes [3], [25]. These methods can be classified into model-based methods and data-driven virtual sensors.

A family of activated sludge models (ASM) along with a revised model for membrane bioreactor (MBR) were introduced to describe the dynamics of the activated sludge process [5], [10], [14]. In real WWTPs, ASM models require calibration for reliable operations as proposed by Guo in [9] who developed a model calibration and validation algorithm using reduced order activated sludge model [18]. Model-based methods provide a good strategy for estimation however their main drawback is the non-accessibility of variables that are important for model calibration and estimation along with their limitations to express the actual process conditions [12], [25]. Consequently, data-driven estimation algorithms can be utilized to overcome the complexity of implementing model-based methods on the non-linear complex process with many non-measurable variables.

Various studies show the effectiveness of neural networks in modeling wastewater treatment process and soft sensing [25]. For example, [2] and [17] used artificial neural network (ANN) for state estimation and effluent prediction. Also, effluent Biochemical Oxygen Demand (BOD) is estimated by a Stacked Autoencoder (SAE) [20]. Nevertheless, the previous learning methods are static under the hypothesis of steady-state and static processes. Real wastewater treatment processes are always dynamic. Therefore, dynamic data-driven soft sensing methods are developed such as Recurrent Neural Network (RNN). However, vanilla RNN has short memory which causes vanishing
gradients. On the other hand, LSTM neural network is proposed in [11] to overcome the vanishing gradient problem and saving long dependency.

Recent works have been done using LSTM in WWTP. In reference [4], an LSTM was proposed to predict effluent concentrations of nitrate and ammonia a few hours ahead. For example in [21], a control strategy is developed to reduce the forecasted ammonia and nitrite concentrations from an LSTM model. Besides estimation and control, LSTM has been utilized for fault detection in the oxidation and nitrification process of WWTP [15]. In addition, the study compared the results with the statistical method and traditional principle component analysis with support vector machine (PCA-SVM). The comparative experiment in [15] showed that the LSTM gives satisfactory performance and has the advantage to handle non-linear dynamic processes.

The lack of an estimation method to monitor the bacteria, MLSS and MLVSS in WWTP and challenges in the model-based methods, motivate to design a data-driven estimation method. A crucial key challenge was the lack of experimental data and thus a Wasserstein generative adversarial network with gradient penalty (WGAN-GP) is designed for synthetic data generation. In addition, an online adaptive long short term memory (LSTM) neural network is proposed to monitor quality variables in WWTP. As a result, WWTP performance can be assessed by measuring the bacterial concentrations at different phases of the treatment process. Also, the operational behavior and solids inventory of the plant are continuously monitored. The proposed method is tested to monitor the MLSS, MLVSS, influent and effluent bacteria cells count in samples obtained from WWTP located at King Abdullah University of Science and Technology (KAUST). The presented paper contains the following contributions:

- Design and tune a Wasserstein generative adversarial network feasible data generation.
- Develop an estimation method for quality variables (bacteria cell counts, MLSS and MLVSS) in WWTP using LSTM.
- Enhance the performance of the LSTM by developing an adaptive LSTM neural network.

This paper is organized as follows: In Section II the problem formulation is described and KAUST WWTP is introduced along with the datasets. Section III shows the synthetic data generation by WGAN-GP and adaptive LSTM. Then, monitoring of KAUST WWTP and the performance of the proposed method are presented in Section IV. Finally, concluding remarks of the proposed methods are presented in Section V.

II. PROBLEM FORMULATION AND PLANT DESCRIPTION

The main objective of this paper is to monitor the quality of wastewater treatment plants in reducing pollutants to purify water. To do so, bacteria cell concentrations, MLSS and MLVSS are estimated to assess the plant quality. In this work, the LSTM neural network is designed to estimate essential quality variables in WWTP. An important key in designing neural network-based estimation is the lack of data. In addition, the difficulties and time-consuming process of counting bacteria cell numbers pose a real challenge. To overcome this problem, the WGAN-GP has been developed to generate realistic training data. Then, an adaptive LSTM neural network is proposed to monitor the bacteria cell concentrations, MLSS and MLVSS in a sliding window. The LSTM hyperparameters are tuned by Bayesian optimization algorithm. In addition, two different datasets for KAUST WWTP are used to test the performance of the proposed method.

A. KAUST wastewater treatment plant

In this work, data was collected from the WWTP located in KAUST. The plant treats domestic wastewater to prevent its discharge to the environment and uses the treated water (effluent) in irrigation. The plant is based on an activated sludge process with aerobic membrane bioreactor. KAUST plant has a restricted effluent quality criteria that must be satisfied. KAUST plant layout is shown in Figure. There are two stream processes operated in KAUST WWTP where the data are collected and used to evaluate the performance of the proposed method.

The plant consists of four tanks: anoxic tank, aerobic tank and two membrane tanks. The wastewater is treated with fine screening and grit removal before entering the equalization tank. The influent wastewater is then pumped into the anoxic tank and discharged from the MBR tanks with an average flow rate 9500 m$^3$/d. Then, the sludge is recycled from the final tank with a flow rate equals to 113 m$^3$/d to the first tank to enhance denitrification of the nitrate. The air is injected in tank 2 to supply sufficient oxygen to the process through a fine air bubble blower. Also, MBR tanks are supplied with oxygen through a coarse air bubble blower for fouling mitigation. The capacity of the plant is 3151 m$^3$ and the volumes of equalization, anoxic, aerobic and membrane tanks are 1787 m$^3$, 181 m$^3$, 388 m$^3$ and 795 m$^3$, respectively.

B. Datasets

In this work, two WWTP datasets are used for data generation and monitoring. The first dataset has daily measurements for four months periods in 2020 collected from KAUST WWTP (dataset$_1$). The second one is influent and effluent concentrations for KAUST WWTP (dataset$_2$). The measurements are monthly sampled data from August 2018 to January 2020. Both datasets variables are on-site measured despite of the bacterial concentration that is obtained in the lab by flow cytometry based on protocols described earlier [23]. Table I shows all measurements collected from the plant with their maximum and minimum values. The max and min values
TABLE I

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Measurement</th>
<th>Max</th>
<th>Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataset1</td>
<td>MLSS\text{aera}</td>
<td>(18600, 14200)</td>
<td>(5300, 9900)</td>
</tr>
<tr>
<td></td>
<td>MLSS\text{MBR1}</td>
<td>(16400, 16900)</td>
<td>(7100, 11300)</td>
</tr>
<tr>
<td></td>
<td>MLSS\text{MBR2}</td>
<td>(16400, 17600)</td>
<td>(7300, 11500)</td>
</tr>
<tr>
<td></td>
<td>DO\text{aera}</td>
<td>(2.13, 2.6)</td>
<td>(2, 2)</td>
</tr>
<tr>
<td></td>
<td>DO\text{MBR1}</td>
<td>(6.72, 6.12)</td>
<td>(0.07, 6)</td>
</tr>
<tr>
<td></td>
<td>DO\text{MBR2}</td>
<td>(6.58, 6.11)</td>
<td>(0.06, 1)</td>
</tr>
<tr>
<td></td>
<td>MLVSS\text{aera}</td>
<td>(10200, 10400)</td>
<td>(3900, 7400)</td>
</tr>
<tr>
<td></td>
<td>MLVSS\text{MBR1}</td>
<td>(12100, 12500)</td>
<td>(5300, 8400)</td>
</tr>
<tr>
<td></td>
<td>MLVSS\text{MBR2}</td>
<td>(12200, 12700)</td>
<td>(5700, 8600)</td>
</tr>
<tr>
<td></td>
<td>Conductivity</td>
<td>(971, 949)</td>
<td>(601, 578.5)</td>
</tr>
<tr>
<td></td>
<td>TDS</td>
<td>(761.5, 659)</td>
<td>(185.5, 337)</td>
</tr>
<tr>
<td></td>
<td>TSS</td>
<td>(2, 2)</td>
<td>(0.987, 1)</td>
</tr>
<tr>
<td></td>
<td>COD</td>
<td>(8.5, 10)</td>
<td>(2, 2)</td>
</tr>
<tr>
<td></td>
<td>Turbidity</td>
<td>(1.285, 1.525)</td>
<td>(0.275, 0.255)</td>
</tr>
<tr>
<td>dataset2</td>
<td>Bacteria cell</td>
<td>(12.6, 9.81)</td>
<td>(11, 7.87)</td>
</tr>
<tr>
<td></td>
<td>Conductivity</td>
<td>(1020, 5409)</td>
<td>(583, 1663)</td>
</tr>
<tr>
<td></td>
<td>TDS</td>
<td>(762, 4188)</td>
<td>(346, 1176)</td>
</tr>
<tr>
<td></td>
<td>TSS</td>
<td>(175, 3)</td>
<td>(34, 1)</td>
</tr>
<tr>
<td></td>
<td>COD</td>
<td>(298, 9)</td>
<td>(137, 4)</td>
</tr>
<tr>
<td></td>
<td>Turbidity</td>
<td>(Na, 2.5)</td>
<td>(Na, 0.29)</td>
</tr>
</tbody>
</table>

of Dataset 1 represent the maximum and minimum of measurements collected from two different streams (e.g. Max(COD) = \(\text{stream}_{\text{ACOD}}\)\text{stream}_{\text{BCOD}}\). Similarly, in dataset 2 where the maximum and minimum represent the influent and effluent flow (e.g. Max(COD) = \(\text{influent}_{\text{COD}}\)\text{effluent}_{\text{COD}}\).

III. METHODS

In this part, the methods used for monitoring WWTP are introduced. Firstly, the features used for monitoring and data preprocessing is described. Then, a synthetic data generation method is designed. Finally, an adaptive LSTM neural network is developed for the monitoring WWTP. Figure 1 represents a schematic diagram describing the proposed method for quality monitoring.

A. Data preprocessing

For developing an efficient and accurate neural network, input features and output data needed to be prepared sufficiently. Firstly, the Pearson correlation coefficients are computed for input features to remove redundant information and simplify computations. Figure 2 shows that TSS and TDS are highly correlated with turbidity and conductivity respectively. In addition, TSS and turbidity can be used interchangeably sometimes to describe the clarity of the water while the conductivity is linearly correlated with TDS [22]. Hence, using all of them can be computationally costly and add no further information. Therefore, TSS and TDS are neglected in the development of the WGAN-GP and LSTM neural networks where the rest of the variables are used to monitor MLSS, MLVSS and bacteria concentrations.

In the training of the WGAN-GP and LSTM networks the input and target variables are normalize between \((-1, 1)\) to obtain an accurate estimation. There are some missing values in the datasets that are interpolated by a spline of second order. Furthermore, the collected data were filtered by a moving average of window size 4 to filter the data and reduce the noise effect.

B. Wasserstein generative adversarial network

In some applications, measurement data is difficult, inefficient or impractical to be collected. Therefore, synthetic data is required to develop and validate estimation methods such as neural networks [6]. In this work, synthetic data generation method is developed based on Generative adversarial network with Wasserstein loss. In [7], two neural networks compete with each other in a min-max game. Firstly, the generator \(G\) estimates a real data distribution \(p_r\) and a discriminator network \(D\) which classifies the data into real or fake data. The generator starts by sampling latent points \(\Gamma \sim p(\cdot)\) to generate fake data and maximize the probability of \(D\) making incorrect classification. However, vanilla GAN minimizes the difference between the real and fake distribution using cross-entropy loss that results in a slow and unstable training process. In addition, learning curves are not correlated to the sample quality and hence WGAN-GP is proposed to help stabilize and improve the training of GAN [1], [8].

WGAN minimizes the Earth Mover Distance (EMD) between \(p_r\) and the generated data distribution \(p_g\) under the optimal critic. The EMD is the minimum cost of transporting mass to convert the data distribution and is mathematically given by:

\[
\text{EMD}(p_r, p_g) = \inf_{\gamma \in \Pi(p_r, p_g)} \mathbb{E}_{x,y \sim \gamma} [||x - y||], \quad (1)
\]

where \(\Pi(p_r, p_g)\) is the set of all joint distributions between \(p_r\) and \(p_g\). \(\gamma\) denotes all possible transport plan. Then, equation (1) is simplified by Kantorovich-Rubinstein duality [24] to:

\[
\text{EMD}(p_r, p_g) = \sup_{||f||_{L1} \leq 1} \mathbb{E}_{x \sim p_r} [f(x)] - \mathbb{E}_{x \sim p_g} [f(x)]. \quad (2)
\]

where \(f\) is a 1-Lipschitz function satisfies the Lipschitz constraint:

\[
|f(x_1) - f(x_2)| \leq |x_1 - x_2|. \quad (3)
\]

WGAN-GP uses gradient penalty on the critic \(D\) output to enforce the Lipschitz constraint. A differentiable function \(f\) is considered as 1-Lipschitz if and only if it has gradients with norm at most 1 everywhere (\(||\nabla f||_2 \leq 1\)) [8]. The WGAN-GP loss function is given as follows:

\[
\min_G \max_D L(D,G) = \mathbb{E}_{\hat{x} \sim p_{\hat{x}}} D(\hat{x}) - \mathbb{E}_{x \sim p_r} D(x) + \lambda \mathbb{E}_{\tilde{x} \sim p_{\tilde{x}}} (||\nabla_{\tilde{x}} D(\tilde{x})||_2 - 1)^2, \quad (4)
\]

where \(\hat{x} \sim p_{\hat{x}}\) is uniformly sampled between \(x\) and \(\hat{x}\) which are pair of points from the \(p_r\) and \(p_g\), respectively. The structure and hyperparameters of the WGAN-GP are chosen based on a grid search algorithm where each
Fig. 1. Neural Network Based Sensing of Bacteria in Wastewater Treatment Plants

Fig. 2. Pearson correlation coefficient of the input measurements is used to monitor MLSS, MLVSS and bacteria cell counts. The upper part represents dataset 1 correlation and the lower part is the dataset 2 correlation. The lower right part is the effluent data correlation and on the lower left part is the correlation coefficients for the influent dataset.

combination is repeated 5 times to reduce the variance. Table. II shows the set of different choices WGAN-GP models evaluated on a held-out validation set. The grid search of the WGAN-GP procedure is shown in Algorithm 1.

In the WGAN-GP, ReLU activation functions and linear output layer are used in $G$ and $D$. The input of the generator is a latent vector sampled from a uniform distribution of size 20. In all experiments, we found out that the penalty coefficient $\lambda = 20$ and ADAM optimizer $[8], [13]$ works well across different datasets. The datasets are divided into 20% validation set and 80% training set. However, batch sizes of 16 and 4 are found to work well for dataset 1 and dataset 2, respectively, considering the sample size of both data-sets. Table. II shows the search space and the best hyperparameter combination.

C. Adaptive Long short term memory

The LSTM neural network was first proposed by [11] to overcome the vanishing gradients problem in simple RNN. In this paper, an adaptive LSTM neural network is proposed to online monitor bacterial concentration, MLSS and MLVSS in KAUST WWTP. This method is adaptive in the sense that it keeps training whenever new targets are estimated as it is described in the testing phase of Algorithm 2.

The LSTM cell uses two kinds of activation functions $tanh$ and $sigmoid$ which output $M_t$ and $H_t$ memory state and hidden state at time step $t$ as follows:

$$H_t = tanh(M_{t-1})O_t, \quad M_t = \sigma(F_tM_{t-1} + I_tC_t),$$

such that $O_t$ is the output gate, $F_t$ is the forget gate, $I_t$ is the input gate and $C_t$ is the candidate. They are defined as:

$$O_t = \sigma(x_tU^O + H_{t-1}W^O), \quad F_t = \sigma(x_tU^F + H_{t-1}W^F),$$

$$I_t = \sigma(x_tU^I + H_{t-1}W^I), \quad C_t = tanh(x_tU^C + H_{t-1}W^C).$$

where the cell receives the input vector $X_t$, hidden state $H_{t-1}$ and memory state $M_{t-1}$ from previous cell at time step $t-1$. $U$ and $W$ are the corresponding wight matrix for the input vector and hidden state respectively. Figure. 3 represent the LSTM unit.

In the LSTM neural network training, the choice of the hyperparameters is challenging and critical to the performance of the network. Consequently, a good combination of hyperparameters can improve the estimation
Algorithm 1 Grid search of the WGAN-GP

Require: $\lambda$ is the gradient penalty weight. $\alpha$ is the learning rate. $m$ is the batch size, $n_{\text{critic}}$ is the number of iterations of the critic per generator iteration and $\eta$ is the possible hyperparameters combination.

Require: $\Omega_0$ is the initial critic parameters. $\theta_0$ is the initial generator’s parameters.

1: for Hyperparameters set $\in \eta$ do
2: while $\theta$ has not converged do
3: for $t = 1, 2, \ldots, n_{\text{critic}}$ do
4: for $j = 1, 2, \ldots, m$ do
5: Sample real data $x \sim p_r$.
6: Sample latent variable $\Gamma \sim p(\gamma)$.
7: Random number $\epsilon \in U[0, 1]$.
8: $\tilde{x} \leftarrow G_\theta(\Gamma)$.
9: $x \leftarrow \epsilon x + (1-\epsilon)\tilde{x}$
10: $L_j \leftarrow D_\Omega(\tilde{x}) - D_\Omega(x) + \lambda(||\nabla_x D_\Omega(\tilde{x})||_2 - 1)^2$
11: end for
12: $\Omega \leftarrow \text{Adam}(\nabla \Omega , \frac{1}{m} \sum_{j=0}^{m} L_j , \Omega , \alpha)$
13: end for
14: Sample a batch from latent variable $\{\Gamma^{(j)}\}_{j=1}^{m} \sim p(\gamma)$.
15: $\theta \leftarrow \text{Adam}(\nabla \theta , \frac{1}{m} \sum_{j=0}^{m} -D_\Omega(G_\theta(\Gamma), \theta, \alpha))$
16: end while
17: end for

Fig. 3. Scheme of the LSTM cell with $\text{sigmoid}$ recurrent activation for each gate and $\text{tanh}$ activation functions.

Algorithm 2 Adaptive online LSTM

Require: $W$ is window size, $J$ is number of epochs $M_{\text{train}}$ is size of training data and $M_{\text{test}}$ is size of testing data.

Require: Training.

1: for epoch $= 1, 2, \ldots, J$ do
2: for $i = 1, 2, \ldots, M_{\text{train}}$ do
3: Predict the output concentrations $\hat{Y}_{\text{train}}$ from input measurements $[X_{t-w}, X_{t}, \ldots, X_{t}^{M_{\text{train}}}].$
4: end for
5: Monitor the training and validation loss for early stopping.
6: end for

Require: Testing.

7: for $i = 1, 2, \ldots, M_{\text{test}}$ do
8: Estimate output concentrations $\hat{Y}_{\text{test}}$ from input measurements $[X_{t-w}, X_{t}, \ldots, X_{t}^{M_{\text{test}}}].$
9: Save $\hat{Y}_{\text{test}}$ as new observed concentration.
10: Train and update the network weights using $[X_{t-w}, X_{t}, \ldots, X_{t}^{M_{\text{test}}}]$ and $\hat{Y}_{\text{test}}.$
11: end for

TABLE II

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>WGAN-GP</th>
<th>LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of layers</td>
<td>1-3</td>
<td>1-2</td>
</tr>
<tr>
<td>Number of neurons</td>
<td>10-30</td>
<td>10-10</td>
</tr>
<tr>
<td>Batch Normalization</td>
<td>(True, False)</td>
<td>(True, False)</td>
</tr>
<tr>
<td>Learning rate</td>
<td>(0.001, 0.0001)</td>
<td>0.0001, 0.00001</td>
</tr>
<tr>
<td>Dropout</td>
<td>-</td>
<td>(0-0.4)</td>
</tr>
<tr>
<td>Number of layers</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Number of neurons</td>
<td>20,20</td>
<td>10,6</td>
</tr>
<tr>
<td>Batch Normalization</td>
<td>(False)</td>
<td>(True)</td>
</tr>
<tr>
<td>Learning rate</td>
<td>(0.0001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Dropout</td>
<td>-</td>
<td>(0)</td>
</tr>
</tbody>
</table>

We tested the performance of the proposed method on two datasets of KAUST WWTP. In order to efficiently train the LSTM network, synthetic training data is generated by WGAN-GP to increase the sample size into 4 multiple of the size of the real dataset. The combined datasets are split into (68%,17%,15%) for training, validation and testing, respectively. Then, an adaptive LSTM is trained using a batch size equals to 1 to adapt the online training where a single batch is the input feature for a lagging window of size 4. The window size is balanced between computational cost and prediction performance. The larger the window size, the more computations are required and may result in overfitting. Also, the process is working dynamically where smaller window size guarantee more accurate estimation. In order to avoid overfitting a stopping criterion is implemented to track the validation loss. Furthermore, a traditional LSTM network is used to compare the performance of the adaptive LSTM.

A. Bacterial concentration monitoring

In this subsection, the results of estimating influent and effluent bacteria cell counts for KAUST WWTP are presented. First, the influent conductivity and COD are used as inputs to the LSTM to measure the influent bacteria cell counts. Then, the effluent bacteria cell counts performance significantly. Therefore, a tuning algorithm based on Bayesian optimization and Keras Tuner library [19] is implemented minimize the loss on a held-out validation set. Table III shows the tuning space and the optimal combination obtained from the Bayesian algorithm. It is worth mentioning that each combination is trained 3 times to obtain low variance results.

IV. KAUST WWTP monitoring

We tested the performance of the proposed method on two datasets of KAUST WWTP. In order to efficiently train the LSTM network, synthetic training data is generated by WGAN-GP to increase the sample size into 4 multiple of the size of the real dataset. The combined datasets are split into (68%,17%,15%) for training, validation and testing, respectively. Then, an adaptive LSTM is trained using a batch size equals to 1 to adapt the online training where a single batch is the input feature for a lagging window of size 4. The window size is balanced between computational cost and prediction performance. The larger the window size, the more computations are required and may result in overfitting. Also, the process is working dynamically where smaller window size guarantee more accurate estimation. In order to avoid overfitting a stopping criterion is implemented to track the validation loss. Furthermore, a traditional LSTM network is used to compare the performance of the adaptive LSTM.

A. Bacterial concentration monitoring

In this subsection, the results of estimating influent and effluent bacteria cell counts for KAUST WWTP are presented. First, the influent conductivity and COD are used as inputs to the LSTM to measure the influent bacteria cell counts. Then, the effluent bacteria cell counts
are estimated from the effluent turbidity, conductivity and COD.

Estimation results of bacteria cell counts are shown in Figure. 4 which presents the LSTM and adaptive LSTM predictions. The adaptive LSTM shows a slight enhancement in the performance. The estimation of influent and effluent bacteria cell counts estimation is in the log scale. The results show that both LSTM networks have obtained a satisfactory performance although the training is implemented on synthetic data. Hence, the WGAN can estimate the distribution of the real data and generate feasible samples for the training of the predictive model.

B. MLSS and MLVSS monitoring

This part focuses on the monitoring of MLSS and MLVSS values for the aeration and MBR tanks of KAUST WWTP. Dataset is obtained for two different streams of the plant. The LSTM network uses dissolved oxygen in all tanks, conductivity, COD and turbidity to predict the concentrations of MLSS and MLVSS for both streams.

The estimation results of MLSS and MLVSS for Stream A and Stream B are presented in Figures. 5 and 6 respectively. The stem plots in each row represent MLSS and MLVSS concentrations in each tank where the first row is the aeration tank, the second row is the MBR tank 1 and the last row is MBR tank 2. The results show that the adaptive LSTM outperforms the traditional LSTM network. Moreover, the improvement is seen clearly in all MLSS and MLVSS graphs where the traditional LSTM could not cope with the changes in the concentrations at some points.

In this work, we evaluated the monitoring performance of the bacteria cell counts, MLSS and MLVSS using two performance indexes. In particular, the Normalized Mean Absolute Errors (NMAE) and Normalized Root Mean Square Error (NRMSE) are computed for both datasets to compare the performance. The NRMSE measures the deviation error in a quadratic sense that is sensitive to outliers whereas, the NMAE is less sensitive to outliers and does not consider the positive and negative discrepancies. Table. III compares the performance of the standard LSTM with the adaptive LSTM based on NMAE and NRMSE. The results show that the adaptive LSTM outperforms the traditional LSTM in all cases. However, there is a slight improvement compared to the MLSS and MLVSS monitoring due to the small testing data.

V. CONCLUSION

In this work, quality monitoring of WWTP based on adaptive neural networks is investigated. The WGAN-GP is used for synthetic data generation and training purposes. Then, the adaptive LSTM is developed to monitor the bacteria cell counts, MLSS and MLVSS in WWTPs. For better performance, WGAN-GP and adaptive LSTM are tuned to find optimal hyperparameters choices. The performance of the proposed method is compared with traditional LSTM using NMAE and NRMSE. The results show that our developed method can monitor water quality in real-time. In addition, an assessment indicator of the cell density of the treatment process is proposed, hence providing a quick assessment of the log removal efficiencies achieved by the WWTP and quality of the final treated effluent. Furthermore, the proposed method indicates the operational behavior and solid inventory in WWTP that controls when to waste or recycle the sludge.

ACKNOWLEDGMENT

The authors would like to acknowledge the assistance from the KAUST Facilities and Management Utilities Team. This work has been supported by the KAUST, Saudi and Center of Excellence for NEOM research at KAUST, Saudi Arabia flagship project research fund (REI/1/4178-03-01).

TABLE III

<table>
<thead>
<tr>
<th>Variable</th>
<th>LSTM</th>
<th>adaptive LSTM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NMAE</td>
<td>NRMSE</td>
</tr>
<tr>
<td>StreamA</td>
<td>0.056</td>
<td>0.081</td>
</tr>
<tr>
<td>StreamB</td>
<td>0.063</td>
<td>0.089</td>
</tr>
<tr>
<td>StreamC</td>
<td>0.071</td>
<td>0.094</td>
</tr>
<tr>
<td>StreamD</td>
<td>0.059</td>
<td>0.091</td>
</tr>
<tr>
<td>StreamE</td>
<td>0.054</td>
<td>0.073</td>
</tr>
<tr>
<td>StreamF</td>
<td>0.062</td>
<td>0.084</td>
</tr>
</tbody>
</table>

REFERENCES

Fig. 4. Influent (left) and effluent (right) bacteria cell counts monitoring in log scale.

Fig. 5. MLSS (left) and MLVSS (right) monitoring results of stream A for LSTM and adaptive LSTM. The rows present the concentrations in each starting from aeration tank, MBR tank 1 and MBR tank 2.


Fig. 6. MLSS (left) and MLVSS (right) monitoring results of stream B for LSTM and adaptive LSTM. The rows present the concentrations in each starting from aeration tank, MBR tank 1 and MBR tank 2.