Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Brewery wastewater treatment plant key-component estimation using moving-window recurrent neural networks

L. Dewasme *

* *Automatic Control Laboratory, University of Mons, 7000 Mons, Belgium (e-mail: Laurent.Dewasme@umons.ac.be).*

Abstract: This work proposes an experimental validation of software sensors for advanced on-line anaerobic digester process monitoring. The considered strategy is based on cheap available measurements (conductivity, temperature, pH, redox potential, etc) to reconstruct key component trajectories such as volatile fatty acid, carbonate and alkalinity concentrations, as well as biogas composition (methane, carbon dioxide, etc). The proposed solution considers a radial basis function artificial neural network ($RBF-ANN$) structure, using data processing (principal component analysis) and an efficient and fast sequential learning algorithm. In order to better reproduce unknown and complex process dynamics, the combination of a moving-window technique with a simple Jordan recurrent ANN structure ($MW-RBF-RNN$) is proposed. Comparative results based on real industrial data illustrate the estimation improvements provided by the $MW-RBF-RNN$ with respect to the classical $RBF-ANN$ structure.

Keywords: Anaerobic digestion; machine learning; software sensor; brewery waste treatment

1. INTRODUCTION

Anaerobic digestion (AD) is a wastewater treatment process that is widely used in the field of brewery since more than half a century (Martins et al. (2018), Tokos and Pintaric (2009)). This success results from economical and ecological advantages, renewing and saving more energy with respect to traditional aerobic treatment which is, indeed, energetically more demanding (McCarty (2001)). However, AD is delicate to control since the operation is sensitive to several chemical parameters as well as the influent content (pH, chemical oxygen demand, suspended solid and volatile fatty acid concentrations, etc) which may be unknown. AD monitoring is therefore crucial to stabilize the process and may require the use of expensive hardware probes. Software sensor design may be an interesting alternative solution, taking advantage of hopefully affordable sensor signals feeding a mathematical model which, most of the time, results from a mechanistic approach with inherent biological interpretability, and interesting predictive capabilities, compensating the lack of frequent measurements of biological variables. Several AD mechanistic models have been proposed in the literature like the IWA ADM1 model (Batstone et al. (2002), Blumensaat and Keller (2005), Alcaraz-Gonzalez and Gonzalez-Alvarez (2007), Lara-Cisneros et al. (2016), Giovannini et al. (2018), Dewasme et al. (2019)]). Mechanistic model-based monitoring strategies present however the drawbacks to require a minimum of on-line information on key-components and to guess an a priori convenient kinetic structure which is not necessarily valid in all process operating scenarios. This paper tackles the problem of AD process key-component level inference from very basic measurements which do not explicitly appear in mechanistic models, as shown in Figure 1. In this particular case, machine learning using artificial neural network (ANN) structures is an appropriate alternative. Successful applications of data-driven strategies to model the complex nonlinear behavior of bioprocesses have been developed during the last decades (Oliveira (2019), Karim et al. (1997), Chen et al. (2000), Renotte et al. (2001), Vande Wouwer et al. (2004), Dewasme et al. (2009)), even if the resulting models often present limited predictive capabilities beyond the studied experimental field.

The current study considers an AD process of brewery wastewater treatment using a plug-flow methanizer (see Figure 1), and aims at designing an advanced monitoring based on intelligent ANN software sensors using basic and cheap available measurements (in the same spirit as in Aceves-Lara et al. (2012) and Charnier et al. (2016)) such as liquid input (F_{in}) and biogas output (F_{out}) flows, pH, temperature (T), conductivity (CON), redox potential (ORP) and total suspended solid concentration (T_{SS}) to estimate key-component levels, that is, volatile fatty acids (VFA), carbonates (CAR) and alkalinity (ALK) concentrations as well as biogas content, i.e. hydrogen sulphide (H_2S), methane (CH_4) and carbon dioxide (CO_2) fractions.

The next section presents the industrial process and available data-sets. A classical radial basis function artificial neural network ($RBF-ANN$) and an original moving-window recurrent NN are presented in section 3. Comparative results of their applications to the available experimental data-sets are proposed in section 4 and conclusions are drawn in section 5.

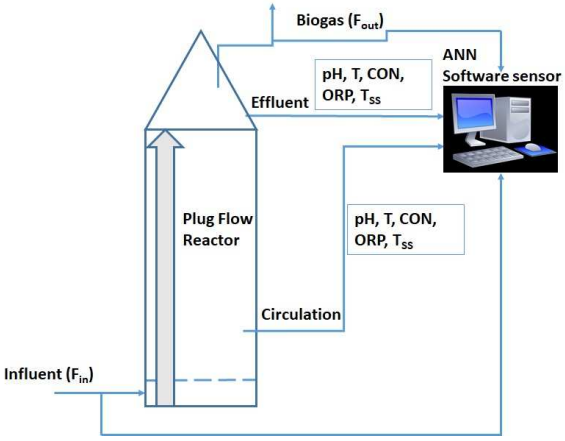

Fig. 1. Schematic view of the considered Plug-Flow Anaerobic Digestion Reactor with the proposed data-driven sensor.

2. PROCESS DESCRIPTION

2.1 Digester monitoring

As illustrated in Figure 1, the considered 60 m^3 anaerobic reactor, located in the brewery "De Halve Maan" in Bruges (Belgium), is fed at the bottom with an influent flow F_{in} recorded every minute. Wastewater chemical composition is measured in two specific locations, the circulation flow and the effluent line. The biogas outflow is also measured as well as its composition by a Combimass gas flow meter and analyser from the Bindergroup measurement and control, chemical parameters by Liquiline Endress and Hauser sensors in the reactor circulation flow and in the effluent, T_{SS} is measured by a Solitax turbidity probe from Hach in mg/L, while VFA, CAR and ALK measurements are achieved analytically from the effluent line thanks to an Anasense device (Applitek) in mg/L.

2.2 Data-sets

A 19-day data-set has been provided by the company Anabel Energy, managing the anaerobic digester monitoring. Since the probes described in subsection 2.1 present different sampling times, a cubic spline (using the function "*interp*1" from MATLAB) is chosen to synchronize the available data at a constant sampling time of 2 hours. The resulting data-set has a total sample number $n_{tot} = 3870$ spread over $n_{in} = 12$ input and $n_{out} = 6$ output signals (described in the next subsection), that is, $n_{samp} = 215$ samples per signal. The proposed data-set partitioning considers the 10 first days for the ANN learning while the 9 remaining days are dedicated to cross-validations. For the sake of confidentiality, the upcoming results are normalized.

3. ARTIFICIAL NEURAL NETWORK STRUCTURE

3.1 A simple radial basis function structure

The selection of a radial basis function (RBF) is motivated by its successful previous applications in biological system state estimation, as, for instance, in Hulhoven et al. (2006) and Dewasme et al. (2009), and the existence of a well-adapted fast learning procedure proceeding in 3 steps described in Vande Wouwer et al. (2004).

Basically, the nonlinear mapping is defined by an input vector $y \in \Re^{1\times n_{in}}$ containing the basic online measurements pH, T, CON, ORP, T_{SS} (in the circulation flow and the effluent line, see Figure 1), inlet and outlet flows F_{in}/F_{out}, and an output vector $x \in \Re^{1\times n_{out}}$ containing the key-component measurements VFA, CAR, ALK, H_2S, CH_4 and CO_2. As stated in Dewasme et al. (2009), it may also be interesting to reduce the number of parameters to accelerate the learning phase, considering a data principal component analysis (PCA, Geladi and Kowalski (1986)) providing a set of uncorrelated linear combinations of the input signals, such that the user may select $n_{PC} < n_{in}$ new inputs while almost conserving the same level of information (at least 95 % in good practice). A new score or principal component matrix S of dimension $n_{samp} \times n_{in}$ is generated by the projection of the measured data matrix Y $(n_{samp} \times n_{in})$ in a new space, using a square orthogonal loading matrix P of dimension $n_{in} \times n_{in}$ as in:

$$S = Y \times P \tag{1}$$

The score matrix is typically generated by iterative partial least squares algorithms, also providing the variances of each component and the corresponding relative level of explanation, i.e. informative content. Table 1 shows the results of the PCA applied to the first 10 days of the available database. Selecting only the first $n_{PC} = \frac{1}{2}n_{in} = 6$ input components is, for instance, enough to conserve more than 95 % of the chosen database informative content.

Table 1. Results of the input signal PCA: PC are sorted by decreasing variance

PC N°	*Variances*	*Explanations*
1	5.2546	47.1888
2	2.1308	19.1355
3	1.3531	12.1516
4	0.9803	8.8032
5	0.7306	6.5609
6	0.2845	2.5547
7	0.1951	1.7517
8	0.1083	0.9725
9	0.0716	0.6433
10	0.0192	0.1725
11	0.0055	0.0493
12	0.0018	0.0158

Considering a simple one-layer RBF structure with k neurons, the i^{th} output therefore reads:

$$\hat{x}_i(t) = \sum_{j=1}^{k} w_{ij} e^{\frac{-\left\|\underline{PC}(t)-\underline{c}_j\right\|^2}{r_j^2}} + b_i \quad i = 1, ..., n_x \tag{2}$$

where $\underline{PC} \in \Re^{1\times n_{PC}}$ is the vector of principal components (linear combinations of the input signals), $W \in \Re^{n_x \times k}$ is the matrix of weights w_{ij}, $\underline{b} \in \Re^{1\times n_x}$ is the bias vector, $C \in \Re^{k\times n_{PC}}$ is the matrix of centroids, where $\underline{c}_j \in \Re^{1\times n_{PC}}$ is the j^{th} raw of C and $\underline{r} \in \Re^{1\times k}$ is the vector of radii.

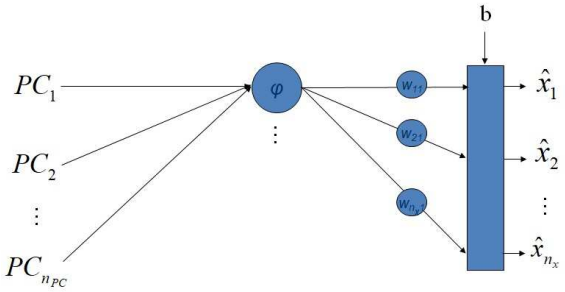

Fig. 2. Radial Basis Function Artificial Neural Network ($RBF-ANN$) structure with PCA.

The number of neurons can unfortunately only be determined by trial and error during the validation steps using, for instance, a simple root mean square criterion (RMS).

The resulting $RBF-ANN$ structure with input preprocessing by PCA is shown in Figure 2 where φ stands for the radial basis function $e^{-(.)}$.

3.2 An original recurrent ANN structure

The $RBF-ANN$ mapping from section 3.1 remains static and is not able to exploit the full dynamical content of the data. It is therefore suggested to consider a receding horizon, also called "moving-window" (MW), of each input/output signal, inserting parallel input/output layers accounting for past samples, introducing a so-called moving-window radial basis function recurrent neural network ($MW-RBF-RNN$) predicting the key-component evolutions in $t+1$ and illustrated in Figure 3. This classical RNN form, also called "Jordan form" (Jordan, 1997), is mostly used in pattern recognition, calculating the most likely output from past output sequences combined to new inputs. In the present study, the same recognition mechanism is applied to predict the bioprocess key-components.

The i^{th} output prediction in $k+1$ becomes:

$$\hat{x}_i(t+1) = \sum_{l=t-h_o+1}^{t} \sum_{j=1}^{k_o} w_{ij,l}^{o} e^{\frac{-\left\|\underline{PC}_o(l)-\underline{c}_{o,j}\right\|^2}{r_{o,j}^2}} + b_{o,i} \quad i=1,...,n_x \tag{3}$$

where the o index stands for the output parameters and variables and the m^{th} component of $\underline{PC}_o$ reads:

$$PC_{o,m}(t) = \sum_{l=t-h+1}^{t} \sum_{j=1}^{k} w_{mj,l}^{o} e^{\frac{-\left\|\underline{PC}(l)-\underline{c}_j\right\|^2}{r_j^2}} + b_m \quad m=1,...,n_{PC_o} \tag{4}$$

It should be noticed that while the numbers of centroids, radii and biases of the first $RBF-ANN$ structure remain unchanged, the number of weights of the input layers is now h times greater. A new PCA is achieved on the $n_{out}=6$ output signals and the results from Table 2 show that $n_{PC_o}=4$ components are sufficient to reproduce more than 95 % of the relative explanation level of the considered part of the database.

4. RESULTS AND DISCUSSION

4.1 Learning algorithm

A three-step learning procedure, described in Vande Wouwer et al. (2004) and Dewasme et al. (2009), is implemented on MATLAB platform to identify the parameters of the classical $RBF-ANN$ and the original $MW-RBF-RNN$ structures, and can be summarized as follows:

- An unsupervised step, using k-means clustering, estimating C and $\underline{r}$;
- A supervised step taking advantage of the radial basis forms of (2), (3) and (4), using fixed values of C and $\underline{r}$ from the first step and estimating W and $\underline{b}$ from a simple linear regression;
- A second supervised step using the previously identified values from the first two steps as initial guesses and solving, numerically, a nonlinear regression problem re-identifying the whole set of parameters in order to cancel possible estimation bias.

This last step consists in solving the following unconstrained nonlinear programming problem:

$$\min_{C,\underline{r},W,\underline{b}} J_{nonlin} = \sum_{t=t_0}^{t_f} \frac{1}{2} \left(\underline{\hat{x}}(t)-\underline{x}(t)\right) Q^{-1} \left(\underline{\hat{x}}(t)-\underline{x}(t)\right)^T \tag{5}$$

where t_f is the experiment final time, Q is the covariance matrix of the measurement errors, assumed diagonal (the measurements present independent noise distributions) and containing the squared maximum values of the outputs, a good practice usually considered when the signal noise variances are assumed to be unknown, in order to uniformly distribute the error independently of the signal orders of magnitude. The Nelder-Mead (simplex) algorithm is chosen to solve (5) using the MATLAB function "fminsearch".

4.2 RBF − ANN validation

The residuals of the direct and cross-validations using the $RBF-ANN$ structure are respectively presented in the left and right sides of Figure 4. J_{cross} is the cross-validation residual and is calculated comparably to J_{nonlin} in (5). In order to avoid overparametrization effects, the selected range of neurons is chosen such that the total number of parameters is always 10 times lower than the total number of samples.

Obviously, the best results are obtained for 8 neurons with an acceptable direct validation residual J_{nonlin} and the

Table 2. Results of the output signal PCA

$PC\ N^o$	$Variances$	$Explanations$
1	4.7880	67.4508
2	1.2706	17.8994
3	0.5495	7.7406
4	0.3084	4.3449
5	0.1426	2.0090
6	0.0330	0.4654

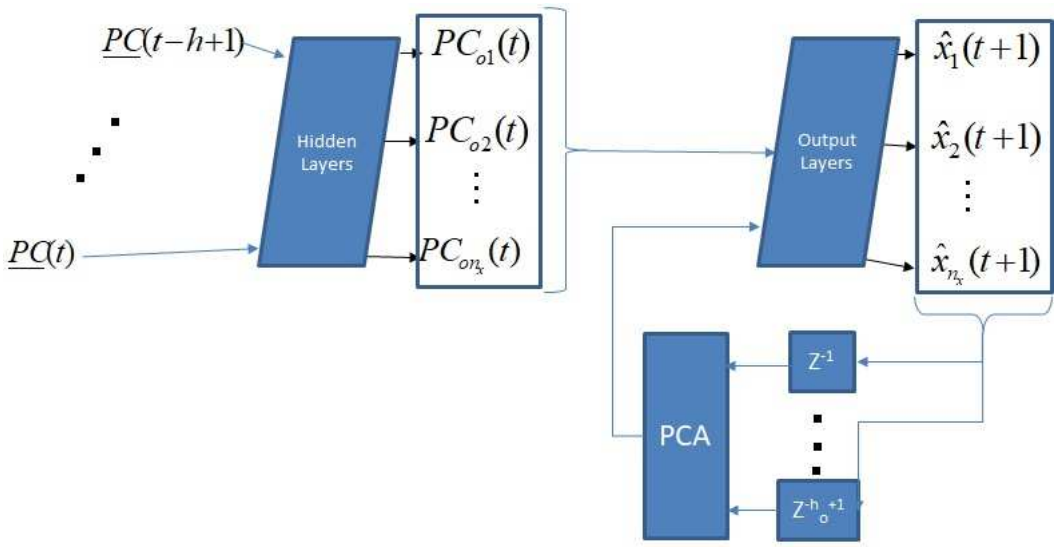

Fig. 3. Moving-Window Radial Basis Function Recurrent Neural Network ($MW-RBF-RNN$) structure with PCA.

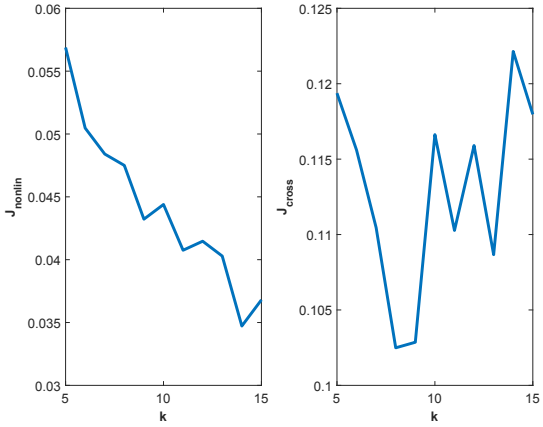

Fig. 4. $RBF-ANN$ direct (left) and cross (right) validation cost function residuals as functions of the number of neurons.

lowest cross-validation residual J_{cross}. The corresponding key-component estimations are shown in Figure 5 where 95% confidence intervals considering a relative standard deviation of 10 % on the measurement errors are drawn to provide a qualitative assessment of the results. The quality of the key-component estimations is, in overall, satisfying, excepted during the last 4 days where VFA, CAR and H_2S present important estimation errors.

4.3 $MW-RBF-RNN$ validation

Inserting moving windows in the NN structure induces an increase of the number of parameters as follows:

$$\begin{aligned} n_{param} =& (h\ k+1)\ n_{PC_o} + (k+1)\ n_{PC} \\ &+ (h_o\ k_o+1)\ n_x + (k_o+1)\ n_{PC_o} \end{aligned} \quad (6)$$

where the terms $(h\ k+1)\ n_{PC_o}$ and $(h_o\ k_o+1)\ n_x$ respectively account for the weights and biases while $(k+1)\ n_{PC}$ and $(k_o+1)\ n_{PC_o}$, respectively for the centroids and radii.

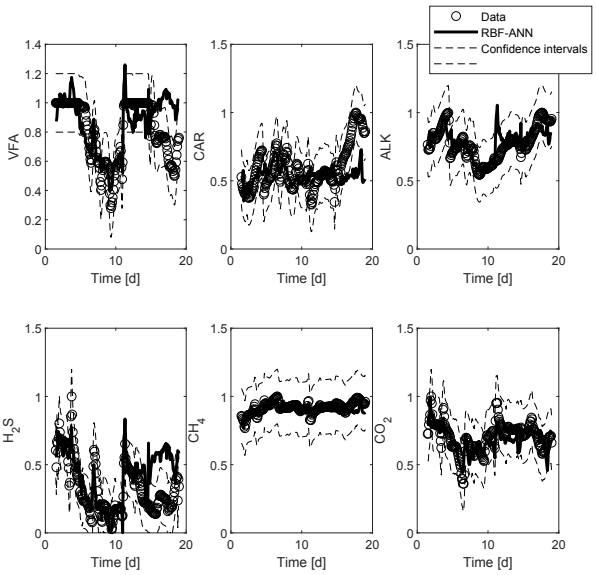

Fig. 5. $RBF-ANN$ cross-validation using one layer of 8 neurons.

However, since the input/output moving-windows respectively require h and h_o times the corresponding input/output data-sets, the total effective number of samples becomes:

$$n_{effsamp} = h\ n_{PC}\ n_{samp} + h_o\ n_{PC_o}\ n_{samp} + n_x\ n_{samp} \quad (7)$$

Table 3 presents the validation results of the $MW-RBF-RNN$ for a selection of parameter combinations, that is, $h, h_o \in \{5,8\}$ and $k, k_o \in \{6,10\}$, such that the considered number of parameters is consistent with respect to the available number of samples following $n_{effsamp} > 10 \times n_{param}$. The new structure direct validation shows better residuals than the simple $RBF-ANN$ from 4 (excepted for one combination) while the cross-validation residuals reach lower values in specific cases, that is, for the combinations $[h, h_o, k, k_o] = [5,5,10,6]$, $[h, h_o, k, k_o] = [8,8,6,6]$

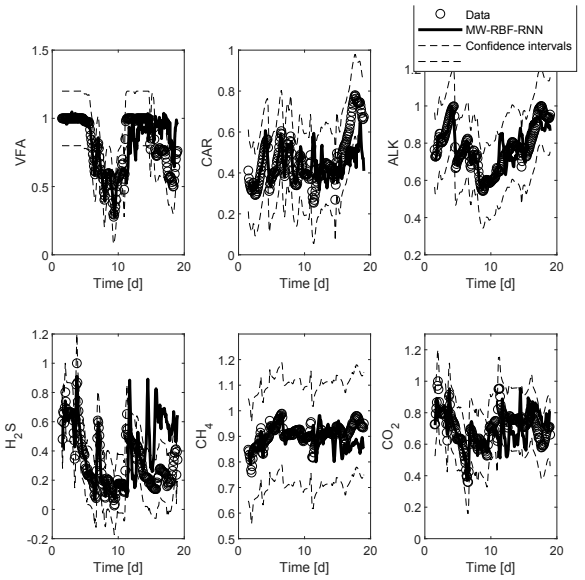

Fig. 6. $MW-RBF-RNN$ cross-validation with $h=5$, $h_o=5$, $k=10$ and $k_o=6$.

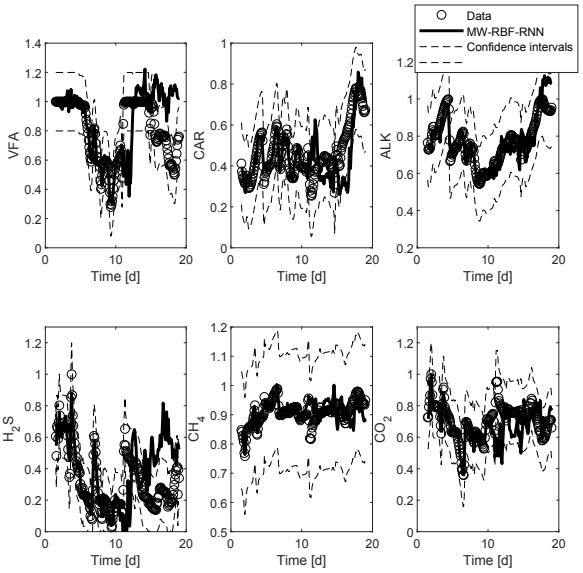

Fig. 7. $MW-RBF-RNN$ cross-validation with $h=8$, $h_o=8$, $k=6$ and $k_o=6$.

and $[h, h_o, k, k_o] = [8, 5, 6, 6]$, which are respectively represented in Figures 6, 7 and 8.

Objectively, it must be pointed out that the quality of CAR estimation is generally poor around the eleventh day and that H_2S presents an abnormally oscillating behavior between days 15 and 19. Anaerobic digestion is a very slow process and the relatively small amount of collected data is likely to limit the performance of the proposed data-driven architecture. Despite these informative limitations, the results are promising since a 10-day data-set is already sufficient to provide an efficient learning with accurate direct validations and, in overall, acceptable cross-validations.

It must also be noticed that the quantitative assessment from Table 3 should not be the only criterion to be taken into account. Figure 9 indeed shows that the combination providing the most faithful reproduction of the process dynamics is $[h, h_o, k, k_o] = [5, 8, 10, 6]$. The higher cross-validation residual ($J_{cross} = 0.1207$) is explained by the worse H_2S estimation.

Table 3. $MW-RBF-RNN$ validation results (horizon length h, number of neurons k, direct and cross-validation residuals J_{nonlin} and J_{cross})

h	h_o	k	k_o	J_{nonlin}	J_{cross}
5	5	6	6	0.0326	0.1914
5	5	6	10	0.0282	0.1420
5	5	10	6	0.0301	0.1034
5	8	6	6	0.0330	0.2348
5	8	10	6	0.0284	0.1207
8	5	6	6	0.0325	0.0925
8	5	6	10	0.0287	0.1261
8	5	10	6	0.0304	0.1367
8	5	10	10	0.0264	0.1903
8	8	6	6	0.0305	0.1055
8	8	6	10	0.1007	0.2164
8	8	10	6	0.0285	0.2534

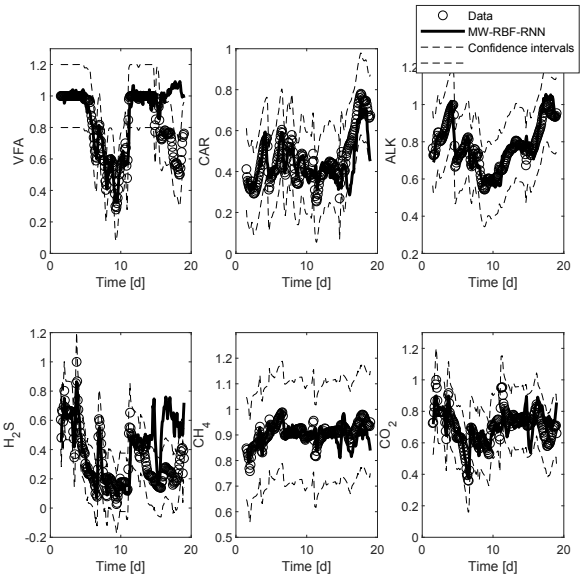

Fig. 8. $MW-RBF-RNN$ cross-validation with $h=8$, $h_o=5$, $k=6$ and $k_o=6$.

5. CONCLUSION

In this work, an original and cheap monitoring device for brewery wastewater digester key components (VFA, CAR, ALK, H_2S, CH_4 and CO_2) is designed. The proposed software sensor resorts to radial basis function neural networks ($RBF-ANN$), allowing a fast machine learning in three short steps. With a 3-week data-set, made available by Anabel Energy (SPRL), validations of basic $RBF-ANN$ are satisfactory but still partially inaccurate. A new contribution using moving-horizons and a recurrent neural network architecture ($MW-RBF-RNN$) is therefore proposed to better capture process dynamics. The performances of the recurrent structure are assessed and the results show better accuracy and tendencies of

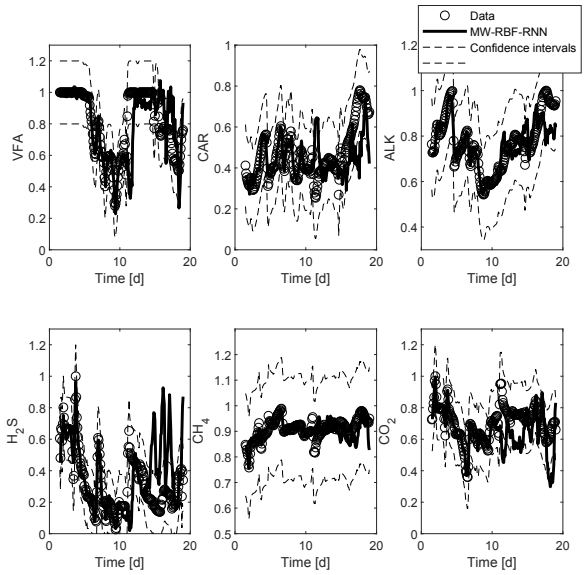

Fig. 9. $MW - RBF - RNN$ cross-validation with $h = 5$, $h_o = 8$, $k = 10$ and $k_o = 6$.

the predictions, which make the $MW - RBF - RNN$ a good candidate for further analysis with richer databases, in view of model-based control design.

ACKNOWLEDGEMENTS

The author sincerely thanks Nikolaas Sobry from Anabel Energy (SPRL) for providing the database used in the present study.

REFERENCES

Aceves-Lara, C.A., Latrille, E., Conte, T., and Steyer, J. (2012). Online estimation of vfa, alkalinity and bicarbonate concentrations by electrical conductivity measurement during anaerobic fermentation. *Water Sci Technol.*, 65(7), 1281–1289.

Alcaraz-Gonzalez, V. and Gonzalez-Alvarez, V. (2007). *Robust nonlinear observers for bioprocesses: Application to wastewater treatment*, volume 361 of *LNCIS*, 119–164. In H. O. Mendez-Acosta, R. Femat and V. Gonzalez-Alvarez (Eds.), Dynamics and Control of Chemical and Biological Processes, springer edition.

Batstone, D., Keller, J., Angelidaki, I., Kalyuzhnyi, S., Pavlostathis, S., Rozzi, A., Sanders, W., Siegrist, H., and Vavilin, V. (2002). The iwa anaerobic digestion model no 1 (adm1). *Water Sci. Technol.*, 45(10), 65–73.

Blumensaat, F. and Keller, J. (2005). Modelling of two-stage anaerobic digestion using the iwa anaerobic digestion model no. 1 (adm1). *Water Research*, 39(1), 171–183.

Charnier, C., Latrille, E., Lardon, L., Miroux, J., and Steyer, J. (2016). Combining ph and electrical conductivity measurements to improve titrimetric methods to determine ammonia nitrogen, volatile fatty acids and inorganic carbon concentrations. *Water Research*, 15(95), 268–279.

Chen, L., Bernard, O., Bastin, G., and Angelov, P. (2000). Hybrid modelling of biotechnological processes using neural networks. *Control Engineering Practice*, 8(7), 821–827.

Dewasme, L., Bogaerts, P., and Vande Wouwer, A. (2009). *Monitoring of bioprocesses: mechanistic and data-driven approaches*, 57–97. Studies in Computational Intelligence, (Computational Intelligent Techniques for Bioprocess Modelling, Supervision and Control, Maria do Carmo Nicoletti, Lakhmi C. Jain, eds.). Springer Verlag.

Dewasme, L., Sbarciog, M., Rocha-Cózatl, E., Haugen, F., and Vande Wouwer, A. (2019). State and unknown input estimation of an anaerobic digestion reactor with experimental validation. *Control Engineering Practice*, 85, 280–289.

Geladi, P. and Kowalski, B. (1986). Partial least squares regression. a tutorial. *Analytica Chimica Acta*, 185, 1–17.

Giovannini, G., Sbarciog, M., Steyer, J., Chamy, R., and Vande Wouwer, A. (2018). On the derivation of a simple dynamic model of anaerobic digestion including the evolution of hydrogen. *Water Research*, 134, 209–225.

Hulhoven, X., Renard, F., Dessoy, S., Dehottay, P., Bogaerts, P., and Vande Wouwer, A. (2006). Monitoring and control of a bioprocess for malaria vaccine production. In *5th IFAC Symposium on Robust Control Design, Toulouse, France*.

Jordan, M. (1997). *Serial Order: A Parallel Distributed Processing Approach*, volume 121 of *Advances in Psychology - Neural-Network Models of Cognition*, 471–495. North-Holland, john w. donahoe, vivian packard dorsel edition.

Karim, M.N., Yoshida, T., Rivera, S.L., Saucedo, V.M., Eikens, B., and Oh, G.S. (1997). Global and local neural network models in biotechnology: Application to different cultivation processes. *Journal of Fermentation and Bioengineering*, 83, 1–11.

Lara-Cisneros, G., Aguilar-Lòpez, R., and Dochain, D. (2016). On-line estimation of vfa concentration in anaerobic digestion via methane outflow rate measurements. *Computers and Chemical Engineering*, 94, 250–256.

Martins, G., Salvador, A., Pereira, L., and Alves, M. (2018). Methane production and conductive materials: A critical review. *Environ. Sci. Technol.*, 52(18), 10241–10253.

McCarty, P.L. (2001). The development of anaerobic treatment and its future. *Water Sci. Technol.*, 44(149-156).

Oliveira, A. (2019). Biotechnology, big data and artificial intelligence. *Biotechnol. J.*, 14(8:e1800613).

Renotte, C., Vande Wouwer, A., Bogaerts, P., and Remy, M. (2001). Neural network applications in non-linear modelling of (bio)chemical processes. *Measurement and Control*, 34(7), 197–201.

Tokos, H. and Pintaric, Z.N. (2009). Synthesis of batch water network for a brewery plant. *Journal of Cleaner Production*, 17, 1466–1479.

Vande Wouwer, A., Renotte, C., and Bogaerts, P. (2004). Biological reaction modeling using radial basis function networks. *Computers and Chemical Engineering*, 28, 2157–2164.