PROCEEDINGS OF THE 32nd European Symposium on Computer Aided Process Engineering
(ESCAPE32), June 12-15, 2022, Toulouse, France
L. Montastruc, S. Negny (Editors)

Soft Sensor of Key Components in Recirculating Aquaculture Systems, using Feedforward Networks

Allyne M. dos Santos[a,*], Espen Karlsen[a], Sigurd Skogestad[a] and Kari J.K. Attramadal[b,c]

[a]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*
[b]*Department of Biotechnology and Food Science, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway*
[c]*Nofitech AS, 7042 Trondheim, Norway*
allyne.dos.santos@ntnu.no

Abstract

Recirculating Aquaculture Systems are known to have its water quality conditions controlled, despite being common to have a simple control structure and a lot of human interaction to achieve that. To avoid long-term exposure to toxic levels of carbon dioxide and ammonia, its concentration needs to be monitored more often than manual measurements are available. In this work, we analyze the multilayer perceptron's ability to monitor water quality components that are important for the development of the fish. This alternative method for monitoring has the potential to complement the current sensor structure and laboratory procedures for manual measurement collection, but more studies need to be done on the type of machine learning model.

Keywords: Soft sensor, Recirculating aquaculture systems (RAS), Feedforward neural networks

1. Introduction

Recirculating Aquaculture Systems (RAS) have two dynamic sub-systems: fish metabolism and water treatment. The water treatment system is responsible by keeping the water quality at high standards, reducing water consumption, and reducing contact with external pathogens (European Market Observatory for Fisheries and Aquaculture Products, 2021).

Regarding the water quality, some components are important to be monitored and controlled due to toxicity, but are hard or not able to be measured continuously, such as ammonia. In addition, it requires either several sensors or a central sensor station with sampling system to get information about the levels of dissolved carbon dioxide and ammonia in all fish tanks and water treatment system. One way of using information of the process to estimate the concentration of these toxic components is by using soft sensors (Fortuna et al., 2007).

The soft sensor technique is a combination of data, for parameter estimation, and process knowledge, for feature selection. The development of the data-driven models can be done, for example, using machine learning models, such as feedforward neural networks (FNN). In this work, the main objective is to apply multilayer perceptron (MLP) (Werbos, 1974), which is a classic type of FNN and a universal approximator (Hornik et al., 1989), as a soft sensor for recirculating aquaculture of Atlantic salmon (*Salmo salar*).

In RAS, even if fish, feed and waste production increase in an exponential way, the goal is that the water quality should be kept stable and good. Measurements of ammonia are performed manually daily, whereas CO_2 is measured continuously at least in one point in the system. However, their levels might vary during the day depending on feeding and other factors, and data about this may be missed. Optimizing the water quality and fish growth would be easier with higher resolution of information. In addition, manual collection and laboratory analyzes are time consuming and therefore costly. Therefore, it is useful to develop an alternative method to monitor these water quality variables.

As the system is assumed to reach a steady state after each day, the water treatment system is approximated to a steady-state model developed in previous work (Dos Santos et al., 2021). The training and validation data are acquired from this model, where some parameters are considered as uncertainties. To improve the soft sensor identification, the uncertainties are changed using latin hypercube sampling (LHS) (Jin et al., 2005), so it contains the operating region, which provides condition for fish optimal growth. After addition of white noise, the data is used to train different MLP configurations for predicting carbon dioxide, ammonia and ammonium concentrations.

2. Process Description

Figure 1 shows a diagram of the RAS this work is foccused on. The process consists of a fish tank, a biofilter, a stripper and an oxygen cone. The model is a simplified version of the process, as it does not consider the effect of the water quality on the fish metabolism, if the conditions are kept within bounds. This assumption is reasonable for each phase of the fish life, which can last from weeks to years depending on a lot of factors. Therefore, this work is only valid for the phase the model represents, which is the smolt phase. This could be easily extended to other phases by changing some parameters in the model, such as the amount of product generated by the fish metabolism per kg of fish feed.

The measurements that are available from sensors or human addition include recirculating volumetric flow rate, q; pH; fish feed rate, F; buffer additions, $\dot{m}_{buffer}$; base additions, $\dot{m}_{base}$; makeup water, q^m; air inlet flow rate, $\dot{m}_{air}$; makeup oxygen, $\dot{m}_{O_2}$; average salinity, S; average temperature, T. More details about the process can be found in Dos Santos et al. (2021).

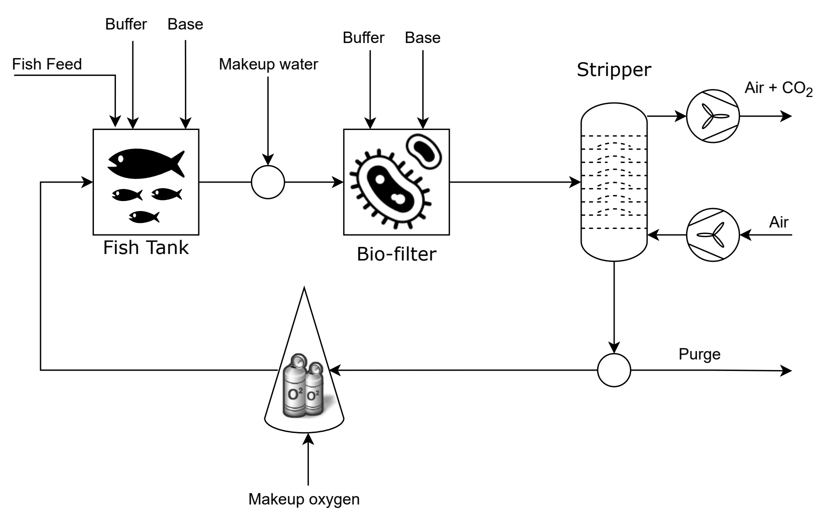


Figure 1: Process diagram of a recirculating aquaculture system

3. Methodology

In order to develop a soft sensor, a standard procedure was followed. Gather data, preprocess it and divide it into training and validation data. After that, fit the model and choose the best model of validation phase. And finally, test it with industrial data.

3.1. Training and Validation Data Acquisition

For the training and validation data acquisition, 6967 steady-state data points were generated within the region described by Table 1, using CasADi v3.5.5 (Andersson et al., 2019) in Python v3.8.8. 5967 of these data points were generated using latin hypercube sampling (LHS) built in `pyDOE` package in Python. LHS is a popular algorithm for planning computer experiments covering the entire range of uncertainties and disturbance in an optimally and distributed way for training. The rest was generated randomly within the same region for validating the soft sensor. After that, 1% white noise was added to both input and output data of both training and validation data.

Table 1: Region of operation

Parameter	**Mean**	**Unit**	**Range**	**Description**
G/L	2.7	-	± 50%	Gas-liquid ratio in equilibrium over the stripper
ξ_B	0.8	-	± 25%	Biofilter efficiency
T	14	°C	± 30%	Average temperature of the system
pH^m	7.0	-	± 10%	pH of the makeup water
$y^{in}_{CO_2}$	4.15e-04	-	± 10%	CO_2 composition in the air inlet
S	15.95	ppt	± 30%	Average salinity of the system
pH^B_{des}	7.2	-	± 1%	Desired pH for the biofilter
q	20	m^3/min	± 50%	Recirculating volumetric flow rate
F	580.6	g/min	+ 50%	Fish feed rate

3.2. Data Preprocessing

Some concentrations are really low in RAS, when comparing with other concentrations. Therefore, it is essential that all data is submitted to preprocessing. As the soft sensor model applied in this work is a deep learning neural network, the normalization of the datasets uses a minmax calculation of the training dataset. After that, the datasets are ready to be used to train and validate the soft sensor.

3.3. Soft Sensor

The feedforward neural network (FNN) architectures were created and optimized using Auto-keras package (Jin et al., 2018), which is a package in Python that automatizes the neural architecture search (NAS) of models supported by another python package named Keras.

In this work, the FNNs are multilayer perceptron (MLP), trained using backpropagation method with a batch size of 500 samples. The objective function of the NAS was the validation loss, the loss function was the mean squared error, and the maximum number of trials is 50. The features were chosen based on knowledge of the process and the available measurements in a real RAS.

After choosing the model with the lowest validation loss, the MLP performance was tested predicting the targets concentration from real data. To avoid breaking the non-disclosure agreement, the collected industrial data was normalized. The performances were compared using the root mean squared error between scaled predicted and scaled data.

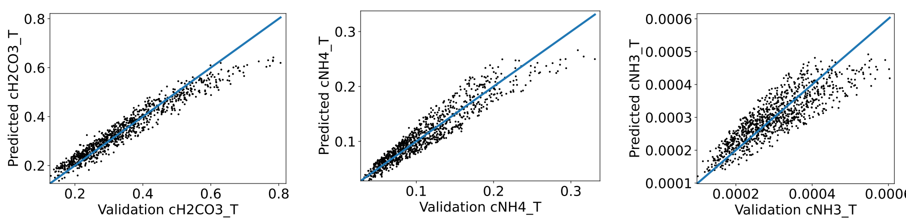
4. Results

The choice of inputs was based on the available measurements from sensors or manual insertion of inlet streams described on the previous section. Three types of models were tested: multiple-input, single-output MLP (MISO-MLP), multiple-input, multiple-output MLP (MIMO-MLP), and a hybrid model, which consists of a MIMO-MLP predicting ammonium and dissolved CO_2 concentrations with ammonia concentration being calculated using the equilibrium equation, see Eq. 1. The inputs of the models were the same for MISO-MLP$_{NH_4^+}$ and MISO-MLP$_{NH_3}$ models: fish feed rate, F, recirculating volumetric flow rate, q, pH in the tank, pH^T. The MIMO-MLP, MISO-MLP$_{H_2CO_3}$ and the hybrid models' features included the same as the previous with addition of a new feature: air inlet flow rate, $\dot{m}_{air}$.

$$c^T_{NH_3} = \frac{K_3(S,T)\, c^T_{NH_4^+}}{c^T_{H^+}} \tag{1}$$

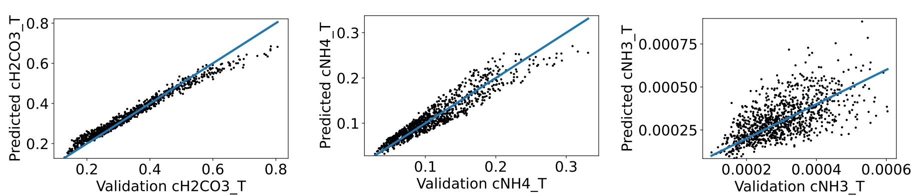
where the equilibrium constant, K_3, is dependent on salinity and temperature, and the concentrations unit is mmol/L.

Figure 2 shows the prediction of the validation data using the MISO-MLP models, Figure 3 shows the results using the hybrid model, and Figure 4 shows the results using the MIMO-MLP model. Comparing Figures 2 and 3, prediction of ammonium and dissolved carbon dioxide were similar, but ammonia predictions are worse using the hybrid model. Comparing Figures 2 and 4, prediction of ammonia is slightly worse on the extremes using MIMO-MLP, and the other predictions were similar.



(a) Prediction of $c^T_{H_2CO_3}$ (b) Prediction of $c^T_{NH_4^+}$ (c) Prediction of $c^T_{NH_3}$

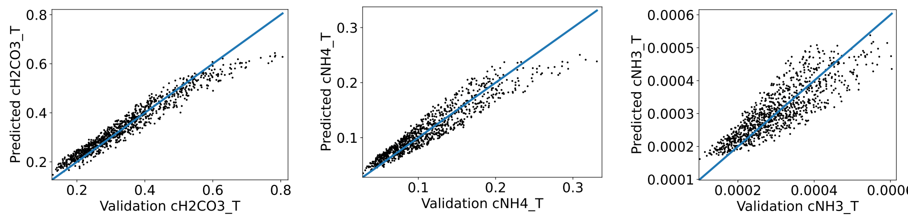
Figure 2: Prediction of validation data using the MISO-MLP models separately



(a) Prediction of $c^T_{H_2CO_3}$ (b) Prediction of $c^T_{NH_4^+}$ (c) Prediction of $c^T_{NH_3}$

Figure 3: Prediction of validation data using the hybrid model

In Table 2, we summarize the performance of the models at the validation phase using the RMSE index. The best MLP architecture for this case study was the MISO-MLP models put together, which gave the lowest final RMSE at the validation phase, and MIMO-NLP model was the second best giving similar, but slightly higher, RMSE.



(a) Prediction of $c^T_{H_2CO_3}$ (b) Prediction of $c^T_{NH_4^+}$ (c) Prediction of $c^T_{NH_3}$

Figure 4: Prediction of validation data using the MIMO-MLP model

Table 2: Summary of the models performance at the validation phase - RMSE index

Output	**MISO-MLPs**	**Hybrid**	**MIMO-MLP**
$c^T_{H_2CO_3}$	0.0645	0.0787	0.0694
$c^T_{NH_4^+}$	0.1204	0.1201	0.1230
$c^T_{NH_3}$	0.1322	0.2611	0.1351
Final	0.1097	0.1720	0.1129

Their architecture of the MLPs are described in Table 3. MISO-MLP$_{H_2CO_3}$ model has two dense hidden layers with 32 nodes each using the rectified linear activation function (ReLU), while the others have one dense hidden layer, but with 64 and 128 nodes using the same activation function on MISO-MLP$_{NH_4^+}$ and MISO-MLP$_{NH_3}$ models, respectively. Note that ammonia concentration turned out to be harder to estimate when compared with ammonium, and possibly the NAS found a flat optimum, which means that probably a MLP with lesser nodes would not improve but would not make the prediction much worse also.

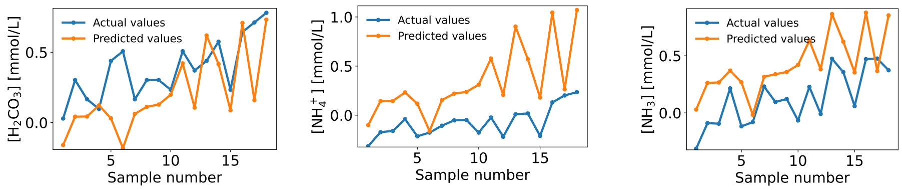
Table 3: Number of nodes in each layer of each MISO-MLP model

Layers	**MISO-MLP$_{H_2CO_3}$**	**MISO-MLP$_{NH_4^+}$**	**MISO-MLP$_{NH_3}$**
Input	4	3	3
Normalization	3	3	3
Dense$_1$	32	64	128
Dense$_2$	32	0	0
Dropout	0	64	128
Output	1	1	1

The prediction of the industrial data using the MISO-MLP models is shown in Figure 5. The predicted values of ammonia concentration seem to be closer to the real data compared with ammonium predictions, which is unexpected. The predicted c$_{H_2CO_3}$ were the same for the first few samples, which means that the model could not extract enough information from the features on those points.

5. Discussion

The prediction of ammonia was revealed to be much harder than of ammonium, resulting in a poor prediction of that variable. This might be due to the noise of the input variables that affects a lot the ammonia concentration, as its magnitude is much smaller. This effect does not disappear after



(a) Prediction of $c^T_{H_2CO_3}$ (b) Prediction of $c^T_{NH_4^+}$ (c) Prediction of $c^T_{NH_3}$

Figure 5: Prediction of industrial data using the MISO-MLP models

preprocessing the data; it expands instead, as the magnitude after normalization is approximately 1000 times higher.

The MISO-MLP models gave the best performance due to different input features to each model, as adding not so important features can make NAS more complex and add a lot of useless cases, as happened with the MIMO-MLP model. This could be solved by increasing the maximum number of trials at the NAS step, but it would take longer, and would still have the possibility of finding different local minimum, for better or worse.

6. Conclusion

The measurement of key waste products are not always easy to collect in real-time or at a required frequency, which reduces the possibility to stabilize and optimize the water quality for the fish. This can be improved by using machine learning models, such as multilayer perceptron.

The MLP models trained in this work are deep neural networks, and its architectures were optimized used an automated neural architecture search and tuning of hyperparameters. Three configurations were compared: MISO-MLPs; hybrid model; and MIMO-MLP. The best configuration was the MISO-MLP models together, although their performance was not so good. This might be due to the possibility of NAS reaching a local minimum or the models could not capture the information it needed, so a different type of model could perform better. A soft sensor using these models for monitoring would perform better than the hybrid model, and the MIMO-MLP model, but would also complement the manual measurements, when estimating dissolved CO_2 and NH_3 concentrations.

References

J. A. Andersson, J. Gillis, G. Horn, J. B. Rawlings, M. Diehl, 2019. CasADi: a software framework for nonlinear optimization and optimal control. Mathematical Programming Computation 11 (1), 1–36.

A. M. Dos Santos, L. F. Bernardino, S. Skogestad, K. J. Attramadal, 2021. Steady-state and Dynamic Modelling of Water Quality in Recirculating Aquaculture Systems. In Progress.

European Market Observatory for Fisheries and Aquaculture Products, 2021. Freshwater aquaculture in the EU. No. March. European Union.

L. Fortuna, S. Graziani, A. Rizzo, M. G. Xibilia, 2007. Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control). Springer-Verlag London.

K. Hornik, M. Stinchcombe, H. White, 1 1989. Multilayer feedforward networks are universal approximators. Neural Networks 2 (5), 359–366.

H. Jin, Q. Song, X. Hu, 6 2018. Auto-Keras: An Efficient Neural Architecture Search System. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1946–1956.
URL `https://arxiv.org/abs/1806.10282v3`

R. Jin, W. Chen, A. Sudjianto, 9 2005. An efficient algorithm for constructing optimal design of computer experiments. Journal of Statistical Planning and Inference 134 (1), 268–287.

P. J. Werbos, 1974. Beyond Regression : New Tools for Prediction and Analysis in the Behavioral Sciences. Ph.D. thesis, Harvard University.