

Forecasting Aquaponic Systems Behaviour With Recurrent Neural Networks Models

Juan Cardenas-Cartagena*, Mohamed Elnourani, and Baltasar Beferull-Lozano

WISNET Center, University of Agder, Norway

Abstract

Aquaponic systems provide a reliable solution to grow vegetables while cultivating fish (or other aquatic organisms) in a controlled environment. The main advantage of these systems compared with traditional soil-based agriculture and aquaculture installations is the ability to produce fish and vegetables with low water consumption. Aquaponics requires a robust control system capable of optimizing fish and plant growth while ensuring a safe operation. Intending to support a control system, this work explores the design process of Deep Learning models based on Recurrent Neural Networks to forecast one hour ahead of pH values in small-scale industrial Aquaponics. This implementation guides us through the machine learning life-cycle with industrial time-series data, i.e. data preprocessing, feature engineering, model architecture selection, training, and validation.

1 Introduction

1.1 Motivation

Aquaponics represents an opportunity to produce food in environments where traditional soil-based agriculture and fish farming have complications, e.g. low temperatures, poor water quality, or hard rocky soil. This system is mainly composed of water-based horticulture, known as hydroponics, aquaculture fish-farming, and bio-filtering with nitrifying bacteria [5]. Each of the subsystems is connected through a recirculating water loop, as illustrated in 1. The idea behind Aquaponics is to treat the ammonia from the fish tanks with the bio-filter and hydroponics by boosting a nitrogen cycle.

*Corresponding Author: juan.d.cardenas@uia.no

Thus, the organisms share the same artificial environment to produce food sustainably [11].

Aquaponics must keep a delicate balance between the needs of different organisms interacting in the system. In other words, plants, fish, and bacteria require different water quality and environmental conditions. Those environmental requirements may conflict with each other. Furthermore, external disturbances in the environment may break the balance inside the system. For example, fish may require a pH between 5 and 10, while plants need values near 6.5, and the bio-filter demands a pH bigger than 7.5 for optimal operation. Additionally, seasonal temperature changes represent disturbances in the system, as they have a considerable effect on the water pH, [10]. Therefore, Aquaponics demands a control system capable of optimizing plants and fish growth while ensuring a healthy interaction between all organisms and robustness against external disturbances.

This control problem motivates the use of a model-based control as a framework to make decisions by looking ahead into the system dynamics. However, the interaction between organisms in Aquaponics is too complex to capture its dynamics with just first-principle models. This problem motivates us to design a data-driven model with recurrent neural architecture to forecast environmental variables in the Aquaponics. In particular, we consider a model for pH as first proof of concept.

1.2 State of the Art

Literature about forecasting variables in Aquaponics can be summarized in empirical and data-driven models. Although there are established models for hydroponics and aquaculture based on biological principles [5, 11], those models do not well-

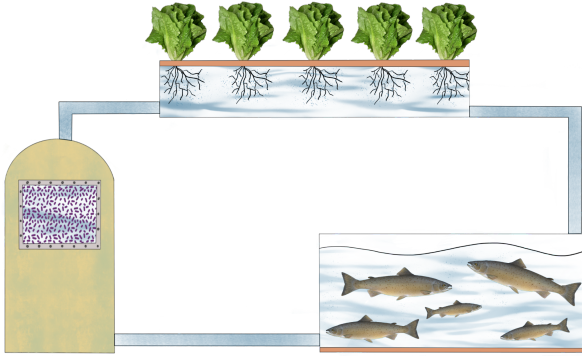


Figure 1: Illustration of Aquaponic system.

characterize the interaction between organisms in real industrial settings. Thus, data-driven modelling becomes a suitable alternative to describe the relationship between variables and predict potential outcomes in the system. Within data-driven models, Deep Neural Networks (DNN) are highlighted due to their flexibility to deal with large-dimensional non-linear systems [6].

As examples of DNN models, consider [4], where a feed-forward neural network is used to forecast pH and electroconductivity in a hydroponics greenhouse. The proposed model forecasts the target variables for 20 minutes ahead with an accuracy of 0.01 pH units and $5 \mu\text{Scm}^{-1}$ respectively. However, this model fails in the presence of control actions, possibly due to the lack of training data with an active controller. A forecasting algorithm for Aquaponics based on empirical biological principle modelling is proposed in [2]. This model is deployed in a small-scale experiment for four weeks. The results suggest that some parameters in the Aquaponics can be forecasted with biological models, e.g. pH in steady-state and total dissolved solids. But, other parameters of interest like pH in transient-state, plant growth, and nitrate concentrations show a considerable error over time.

The information coming from sensors and actuators in Aquaponics has the structure of a time-series. In this case, Recurrent Neural Networks (RNN) provides an architecture where their state outputs are functions of previous states, making them suitable for processing time-series [6]. Although the use of RNN models in Aquaponics is limited, we may consider the case of Aquaculture systems. For example, an RNN model to forecast

dissolved oxygen (DO) for crabs farming is proposed in [8]. The model uses temperature, pH, turbidity, ammonia concentration, and previous DO values as features to predict the DO behaviour for one day ahead. Similar models are proposed in [7, 9] to forecast pH along with other water quality parameters. In these projects, we remark the workflow to pre-processing data, so that there the input variables are relevant to the outputs, and their values are appropriate for the model, i.e. the time-series are smooth and normalized.

Those projects represent the first proof of concept of forecasting algorithms into small-scale Aquaponics and industrial aquaculture via DNN. Their results can be extended into an industrial setting where challenges such as dealing with noisy and incomplete data with unexpected perturbations arise. Hence, this work considers the design of a proof-of-concept for a deep learning model to forecast the dynamics of pH in an industrial Aquaponics setting.

1.3 Contributions

This study constitutes a proof-of-concept for estimation and prediction variables in the Aquaponic via RNN models. For this purpose, the gated recurrent units (GRU) and Long short-term memory (LSTM) architectures are considered to forecast pH in industrial Aquaponics for one hour ahead. The contributions for this project are:

- Implementation of a pre-processing data methodology for the time-series: feature selection, normalization, time-alignment, smoothing, and splitting.
- Design of RNN models under the LSTM and GRU architectures to model pH in Aquaponics.
- Train and validate the models with data from 17 days harvesting period. Use different number of parameters for comparison purposes.
- Compare the results with a multi-layer perceptron network (MLP) as baseline model.

2 Methods

The design workflow for the proposed models is discussed in this section, from the data pre-processing to DNN architecture.

2.1 Preprocessing

As the paper deals with multiple time-series data, then the sequence structure between values is a key factor in the analysis. Furthermore, data in Aquaponics comes with noise from sensors, perturbations in the system, different ranges, and poor quality features with irrelevant samples. Thus, let us discuss how to deal with those issues in the dataset.

2.1.1 Feature selection and data cleaning

As the first step to clean and prepare the dataset, a time frame is selected when there is enough relevant data for training, in this case, from April 1st to 17th, 2021. Then, features with corrupt or unnecessary data are dropout, i.e. data with high noise, a considerable amount of missing values, unlabelled features, or constant time series. For example, signals such as pump set-points, PID controllers, weight cells, and light sensors are discarded. The next step consists of denoising the time series and removing outliers with low-pass filtering, as shown in Fig. 2. After the denoising process, the signal shows a periodic behaviour with some disturbances in the 7th, 9th, and 12th days. Those disturbances are due to harvesting events and the use of regularization substances for water quality.

To reduce the dataset size, the features selection step is based on threshold in the absolute value Pearson correlation coefficient,

$$|r_{x,y}| = \left| \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right|,$$

where $\{(x_i, y_i)\}_{i=1}^n$ is a pair of features, \bar{x}, \bar{y} are their respective sample-mean. in Fig. 3 is a heatmap of $|r_{x,y}|$ for each pair of features in the cleaned dataset. The values near to 1 for the variable 4 (pH in water tanks) are selected as features for the models. The selected features in this process are: pH, dissolved oxygen percentage, water flow, light, and calcium in the water tanks.

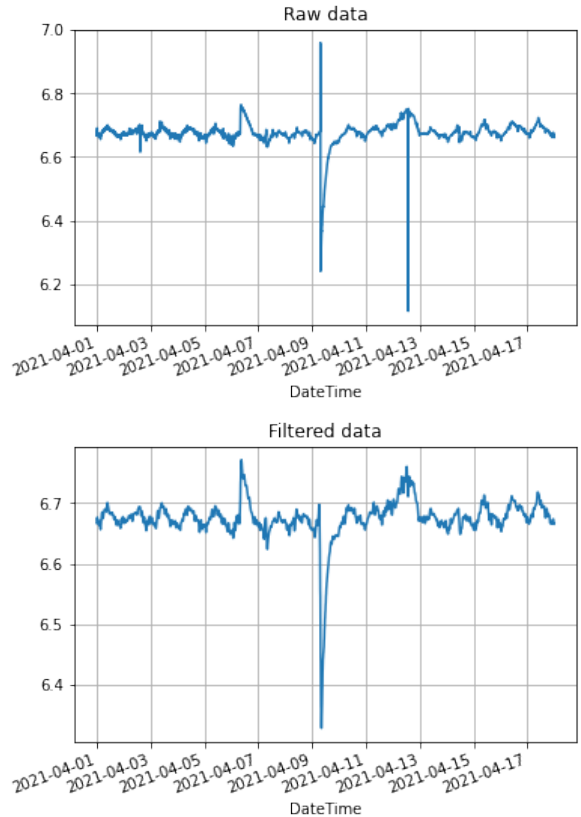


Figure 2: Comparison between raw Ph data with outliers and after filtering

2.1.2 Dataset splitting and Normalization

After cleaning the dataset, the data is split into 70% training, 20% validation and 10% testing out of 28636 samples. Then, it is normalized with a min-max normalization approach

$$z_i = \frac{x_i - x_{min}}{x_{max} - x_{min}}$$

where x_{max}, x_{min} are the maximum and minimum feature values in the overall dataset. Fig. 4 shows the normalized data distribution for the selected features in the training dataset. Some features have different ranges than (0, 1) since the normalization was taken from the overall dataset, so the remaining validation and testing sets have values out of the training one. After normalizing the dataset, we construct feature packages of 3-hours of data per 1-hour of future pH values for training,

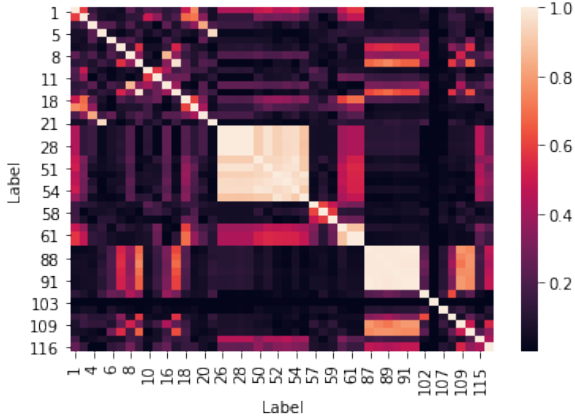


Figure 3: Heatmap of Pearson correlation coefficient absolute value for features

where 1 hour is approximately 65 values. Each feature package contains sequences of the 17 variables selected above with the correlation coefficients.

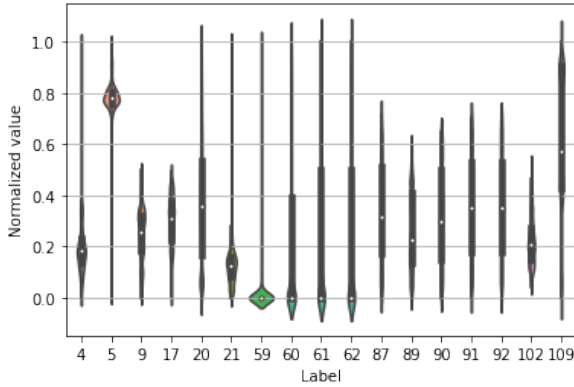


Figure 4: Normalized data in training dataset

2.2 Model Architecture

This project considers the LSTM and GRU architecture to forecast 1-hour ahead of pH. The RNN models were programmed with TensorFlow [1]. The number of layers and neurons for each model was chosen by experimentation. Furthermore, each model had a maximum of 20 epochs for training with a batch size between 50 and 100. The models utilize a Nadam optimizer [3], i.e. an Adam optimizer with momentum. And the loss function

is the Mean Squared Error (MSE),

$$L_{MSE}(z_i, \hat{z}_i) = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2$$

where z_i is the normalized prediction and \hat{z}_i is the normalized true value. Furthermore, the Mean Absolute Error (MAE) is considered as an additional metric for the experiments,

$$L_{MAE}(z_i, \hat{z}_i) = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i|.$$

Meanwhile, the MSE metric penalizes large errors on training, the MAE metric provides a notion of small errors in the forecasting.

Additionally, two callbacks are considered in the models: early stop, to stop training when the loss stops improving; and learning rate scheduler, to decay the learning rate parameter at each epoch. Finally, drop-out layers are considered to reduce the over-fitting in training.

3 Simulation

This section shows the results of the experiments with real data from small-scale industrial Aquaponics and the RNN models.

3.1 pH Forecasting

The pH signal from the sampled period has outliers and an unexpected perturbation on the ninth day of experimentation. However, the variation range is small, i.e. from 6.3 to 6.8. The correlation process suggests a strong relation between pH and the pump motor in the system. The filtering process effectively removed outliers without substantial modifications in the signal. Table 1 summarizes the results for the training and validation process in the baseline and RNN models.

In this case, the LSTM and GRU models have similar results for practical purposes. In particular, the LSTM model has an MSE of $1.56e - 3$; meanwhile, the GRU has a $1.53e - 3$ MSE. Additionally, the both models require similar time to train as it has more parameters than the baseline. Although the evolution of pH over time is slow, both RNN models can forecast with considerable accuracy 1-hour ahead. The forecasting results for 1-hour forecasting are in Fig. 5.

Metric	MLP	LSTM	GRU
MSE Eval.	6.01e-3	1.56e-3	1.53e-3
MAE Eval.	7.21e-2	3.28e-3	3.25e-3
Parameters	6.93e5	2.79e6	3.03e6
Train epochs	3	7	9
Time per epoch [s]	2	29	34

Table 1: Model evaluation metrics

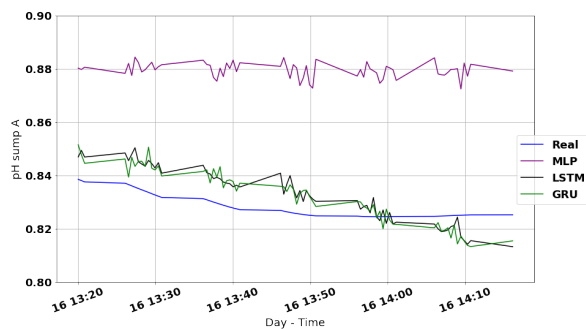


Figure 5: Comparison between different models - Normalized results

4 Conclusion

This paper explores the usage of different RNN models to forecast pH in industrial Aquaponics. This proof-of-concept shows the challenges involved in implementing machine learning models in real-life scenarios. For instance, pre-processing and feature engineering become the most work-expensive steps into this machine learning workflow. In these steps, the main task is to preserve the signal’s behaviour while removing noise and outliers. Additionally, a feature selection process based on a correlation index is utilized to highlight significant inputs in the dataset. Although correlation does not imply causation, the correlation index gives us a relationship notion between variables, especially when describing the behaviour in an artificial environment as Aquaponics.

Concerning the DNN models design, the LSTM and GRU architectures outperform the baseline model considered in this work. Although both models require more trainable parameters, the learning process converges rapidly, and their forecasting results are similar for practical purposes. Most importantly, the RNN models can closely follow the behaviour of a slow dynamics variable as

the pH, in contrast to the baseline model that computes the variable as a constant signal.

The next proof-of-concept consists of implementing a simpler filtering algorithm to reduce noise and outliers in the signals. Moreover, including expert knowledge in the feature selection step to reduce dimensionality while providing some notion of attention in the model. In addition, other architectures, such as auto-encoders and attention mechanisms, may be considered. Finally, the main target from this paper is a model-based controller for Aquaponics. Hence, it is required to research safe operation guarantees from the resulting data-driven model.

Acknowledgement

We thank the INDURB project partners, The Norwegian Institute of Bioeconomy Research (Nibio), Feedback Aquaculture, and The Norwegian Institute for Water Research (NIVA) for supporting this work. This project is funded by the IKTPLUSS INDURB grant 270730/O70 from the Research Council of Norway.

References

- [1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL <https://www.tensorflow.org/>. Software available from tensorflow.org.
- [2] A. Ahmed, S. Zulfiqar, A. Ghandar, Y. Chen, M. Hanai, and G. Theodoropoulos. Digital Twin Technology for Aquaponics: Towards Optimizing Food Production with Dynamic Data Driven Application Systems. In G. Tan, A. Lehmann, Y. M. Teo, and W. Cai, editors, *Methods and Applications for Model-*

- ing and Simulation of Complex Systems*, volume 1094, pages 3–14. Springer Singapore, 2019. ISBN 9789811510779 9789811510786. doi: 10.1007/978-981-15-1078-6_1.
- [3] T. Dozat. Incorporating nesterov momentum into adam. In *The International Conference on Learning Representations*, 2016.
- [4] K. P. Ferentinos and L. D. Albright. Predictive neural network modeling of ph and electrical conductivity in deep-thought hydroponics. *Transactions of the ASAE*, 45:10, 2002. doi: 10.13031/2013.11412.
- [5] S. Goddek, A. Joyce, B. Kotzen, and G. M. Burnell, editors. *Aquaponics Food Production Systems: Combined Aquaculture and Hydroponic Production Technologies for the Future*. Springer International Publishing, Cham, 2019. ISBN 978-3-030-15942-9 978-3-030-15943-6. doi: 10.1007/978-3-030-15943-6.
- [6] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts, 2016. ISBN 978-0-262-03561-3. doi: 10.5555/3086952.
- [7] Z. Hu, Y. Zhang, Y. Zhao, M. Xie, J. Zhong, Z. Tu, and J. Liu. A water quality prediction method based on the deep LSTM network considering correlation in smart mariculture. 19(6):1420, 2019. ISSN 1424-8220. doi: 10.3390/s19061420. URL <https://www.mdpi.com/1424-8220/19/6/1420>.
- [8] W. Li, H. Wu, N. Zhu, Y. Jiang, J. Tan, and Y. Guo. Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Information Processing in Agriculture*, Feb. 2020. ISSN 22143173. doi: 10.1016/j.inpa.2020.02.002.
- [9] J. Liu, C. Yu, Z. Hu, Y. Zhao, Y. Bai, M. Xie, and J. Luo. Accurate prediction scheme of water quality in smart mariculture with deep bi-s-SRU learning network. 8:24784–24798, 2020. ISSN 2169-3536. doi: 10.1109/ACCESS.2020.2971253. URL <https://ieeexplore.ieee.org/document/8979341/>.
- [10] R. Sallenave. Important water quality parameters in aquaponics systems. Technical report, College of Agricultural, Consumer and Environmental Science, New Mexico State University, 2016. URL https://aces.nmsu.edu/pubs/_circulars/CR680.pdf.
- [11] M. Timmons, T. Guerdat, and B. Vinci. *Recirculating Aquaculture*. Ithaca Publishing Company LLC, 4th edition, 2018. ISBN 978-0971264670.