



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2022

THREE ARTICLES ON THE APPLICATION OF MACHINE LEARNING AND ARTIFICIAL NEURAL NETWORKS TO ANALYZE AND SIMULATE DESIGNED ENVIRONMENTAL PROCESSES

McCORMICK Mark William

McCORMICK Mark William, 2022, THREE ARTICLES ON THE APPLICATION OF MACHINE
LEARNING AND ARTIFICIAL NEURAL NETWORKS TO ANALYZE AND SIMULATE DESIGNED
ENVIRONMENTAL PROCESSES

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_50C59F47AB5A9

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES
DÉPARTEMENT DES SYSTÈMES D'INFORMATION

**THREE ARTICLES ON THE APPLICATION OF MACHINE LEARNING AND ARTIFICIAL
NEURAL NETWORKS TO ANALYZE AND SIMULATE DESIGNED
ENVIRONMENTAL PROCESSES**

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de
Docteur ès Sciences en systèmes d'information

par

Mark William McCORMICK

Directeur de thèse
Prof. Alessandro E.P. Villa

Jury

Prof. Rafael Lalive, Président
Prof. Solange Ghernaouti, experte interne
Prof. Antonio J. Pons Rivero, expert externe

LAUSANNE
2022

IMPRIMATUR

Sans se prononcer sur les opinions de l'auteur, la Faculté des Hautes Etudes Commerciales de l'Université de Lausanne autorise l'impression de la thèse de Monsieur Mark William McCORMICK, titulaire d'un bachelor en biochimie de l'Université de Californie, Davis, et titulaire d'un master en ingénierie de l'environnement de l'Université de l'état de Pennsylvanie, en vue de l'obtention du grade de docteur ès Sciences en systèmes d'information.

La thèse est intitulée :

**THREE ARTICLES ON THE APPLICATION OF MACHINE LEARNING
AND ARTIFICIAL NEURAL NETWORKS TO ANALYZE AND SIMULATE
DESIGNED ENVIRONMENTAL PROCESSES**

Lausanne, le 07 septembre 2022

La Doyenne



Marianne SCHMID MAST



Members of the thesis committee

Thesis supervisor:

Professor Alessandro E.P. Villa

Faculty of Business and Economics, Department of Information Systems, University of Lausanne

Internal expert:

Professor Solange Ghernaouti

Faculty of Business and Economics, Department of Information Systems, University of Lausanne

External expert :

Assistant Professor Antonio J. Pons Rivero,

Universitat Politècnica de Catalunya, BarcelonaTech

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

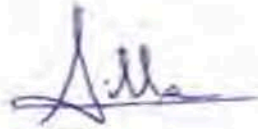
I hereby certify that I have examined the doctoral thesis of

Mark William MC CORMICK

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: _____



Date: _____

5 juillet 2022

Prof. Alessandro VILLA
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

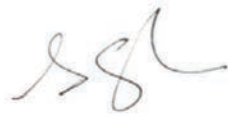
PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Mark William MC CORMICK

and have found it to meet the requirements for a doctoral thesis.

All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.



Signature:



Date 5 juillet 2022

Prof. Solange GHERNAOUTI
Internal member of the doctoral committee

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Mark William MC CORMICK

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

A Pons Rivero

Signature: _____

Date: 11/07/2022

Prof. Antonio J. PONS RIVERO
External member of the doctoral committee

Declaration

I, Mark William McCormick, hereby confirm that there are no known conflicts of interest associated with this manuscript and its three articles. In particular, there has been no financial support for this work that could have influenced and/or biased its outcome. I also confirm that I have given due consideration to the protection of intellectual property associated with this manuscript and its three articles, and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing, I confirm that I have followed the regulations concerning intellectual property of our institutions and the corresponding scientific journals and conferences. I further confirm that any aspect of the work covered in this manuscript that has involved human patients has been conducted with the ethical approval of all relevant bodies and that such approvals are acknowledged within the manuscript.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Alessandro E.P. Villa for guiding me through the long journey from producer of descriptive statistics to modeler of large datasets and latent data structures, and for providing constant learning, technical, and administrative support. He was a pioneer in neural networks 30 years ago and is a lighthouse today in the vast and still largely uncharted ocean of machine learning possibilities.

I would also like to thank Professor Solange Ghernaoui and Professor Antonio Pons Rivero for having accepted to sit on my committee. They provided relevant and valuable criticism that substantially enhanced the coherence of my manuscript.

Thank you to my parents for pointing to the right path in life.

I would like to thank Marie and our children Amélie and Ysaline for contributing to the culture of unsupervised learning that keeps the horizons open and attractive.

The data made this thesis possible. All the acquired data used were generated while I worked on projects conducted independently of the thesis at the Haute école d'ingénierie et de gestion du canton de Vaud (HEIG-VD, Yverdon-les-Bains, CH). I am grateful to all my co-workers and especially to the following organizations and people:

Biomass torrefaction: financed by the canton of Vaud and SCCER Biosweet. The project was initiated and led by Professor Jean-Bernard Michel.

Anaerobic digestion of food wastes: financed by EU-FP7-SME (ORION, n° 282693) and Innosuisse (n° 32609). François-Régis Mahrer designed the anaerobic digester and initiated the project.

Upflow anaerobic filter: financed by SCCER Biosweet and SATW grant n° 2020-001 (initiated and led by me).

Nous sommes faits pour être libres

Nous sommes faits pour être heureux

(Inspiré par Louis Aragon)

Abstract

This thesis consists of three scientific publications that use machine learning methods to understand industrial processes that were designed to improve environmental quality. My work contributes to efforts to make numerical models for simulation and process control that can be applied to improve equipment design and process performance. The personal gain achieved was acquisition of the knowledge and practical skills needed to make better quantitative data analysis and predictive models.

Chapter 1 reports the analysis of data obtained from continuous torrefaction trials conducted in an integrated pilot plant. The purpose of the work was to gain insights into the process behavior that might be useful in the conception of a process controller that aims to obtain a product having a high and constant calorific value from heterogeneous raw biomass having a low and fluctuating calorific value. Principal components analysis of on-line operational data showed that the first 10 principal components account for only 75% of the total variance in the system. This indicates that the data cannot be summarized by only a few latent variables. The Hotelling's t-squared statistic and the first principal component scores were analyzed to identify the periods of process instability and then to identify the disturbance and the manipulated parameters that contribute to the variance during these periods. A Partial Least Squares regression model was developed to map the relation between principal components and the heating value of the torrefied product. The presence of significant and persistent correlations between changes in the values of two parameters, was assessed by determining the correlations at different lag times.

Chapter 2 reports the development of artificial neural network predictive models of data acquired during 12 months of operation of an anaerobic digester treating food wastes. The purpose of the work was to make a predictive model to map process performance, evaluated in terms of CH₄ production, to measurable operational parameters. The models were then assessed for use as decision tools to control the digester feeding rate. Three separate model architectures were evaluated: 1) Long short-term memory (LSTM) 2) Convolutional (CNN) and 3) a hybrid model featuring parallel data flow through LSTM and CNN layers. The LSTM model was found to have the lowest RMSE between predicted and measured biogas production. However, the 1-layer CNN proved to more accurately make correct feeding decisions (58%) which was slightly lower than the rate obtained by skilled operators (63%).

Chapter 3 reports the development of a problem-solving approach and a simulation tool for use in designing upflow anaerobic filters (UAF) for wastewater treatment. The multi-layer perceptron (MLP) artificial neural network model was found to be more accurate than a 4th degree polynomial model and more appropriate than the LSTM model. The Root Mean Squared Error (RMSE) of the predicted responses on unseen test data was 10.1% of the range of the true response values and the coefficient of determination (R^2) was 0.66. The MLP model was then used to run simulations of different design values of the UAF packing diameter, packing material type and the bioreactor height to diameter ratio. The ranked results of simulation were used to select the optimal design values.

Résumé

Cette thèse se compose de trois publications scientifiques qui utilisent des méthodes d'apprentissage automatique pour comprendre des processus industriels qui ont été conçus pour améliorer la qualité de l'environnement. Mon travail contribue aux efforts visant à créer des modèles numériques pour la simulation et le contrôle de processus. Les modèles peuvent être appliqués pour améliorer la conception d'équipements et les performances de processus. Le gain personnel réalisé a été l'acquisition des connaissances et des compétences pratiques nécessaires pour mieux analyser des données quantitatives et mieux concevoir des modèles prédictifs.

Le chapitre 1 rapporte l'analyse des données obtenues à partir d'essais de torréfaction en continu menés dans une usine pilote intégrée. Le but du travail était de mieux comprendre le comportement du procédé et d'extraire des informations qui pourraient être utiles dans la conception d'un contrôleur de procédé visant à obtenir un produit ayant une valeur calorifique élevée et constante à partir de biomasse brute hétérogène ayant une valeur calorifique faible et fluctuante. L'analyse en composantes principales des données opérationnelles en ligne a montré que les 10 premières composantes principales ne représentent que 75 % de la variance totale du système. Cela indique que les données ne peuvent pas être résumées par seulement quelques variables latentes. La statistique t-carré de Hotelling et les scores de la première composante principale ont été analysés pour identifier les périodes d'instabilité du processus, puis pour identifier les paramètres de perturbation et les paramètres manipulés qui contribuent à la variance au cours de ces périodes. Un modèle de régression des moindres carrés partiels a été développé pour décrire la relation entre les composants principaux et le pouvoir calorifique du produit torréfié. La présence de corrélations significatives et persistantes entre les changements des valeurs de deux paramètres a été évaluée en déterminant les corrélations à différents temps de latence.

Le chapitre 2 rapporte le développement de modèles prédictifs à réseaux de neurones artificiels entraînés avec des données acquises pendant 12 mois de fonctionnement d'un digesteur anaérobie traitant des déchets alimentaires. L'objectif des travaux était de réaliser un modèle prédictif permettant de décrire la relation entre les performances du procédé, évaluées en termes de production de CH_4 , et des paramètres opérationnels mesurables. Les modèles ont ensuite été évalués pour être utilisés comme outils de décision pour contrôler le taux d'alimentation du digesteur. Trois architectures de modèles distinctes ont été évaluées : 1) Long short-term memory (LSTM) 2) convolution à 1 couche (1-D CNN) et 3) un modèle hybride présentant un flux de données parallèle à travers les couches LSTM et 1-D CNN. Le modèle LSTM s'est avéré avoir le RMSE le plus bas entre la prédiction de la production de biogaz et la production réellement mesurée. Cependant, le 1-D CNN s'est avéré plus précis comme aide à la prise de décision d'alimentation du digesteur. Le taux de décisions correctes était de 58%, ce qui était légèrement inférieur au taux obtenu par des opérateurs qualifiés (63%).

Le chapitre 3 rapporte le développement d'une approche de résolution du problème de développement d'équipements et un outil de simulation à utiliser dans la conception de filtres anaérobies à flux ascendant (UAF) pour le traitement des eaux usées. Le perceptron multicouche

(MLP) s'est avéré être plus précis qu'un modèle polynomial du 4e degré et plus approprié que le modèle LSTM. L'erreur quadratique moyenne (RMSE) des réponses prédites sur les données de validation était de 10,1% de la plage des valeurs de réponse réelles et le coefficient de détermination (R^2) était de 0,66. Le modèle MLP a ensuite été utilisé pour exécuter des simulations de différentes valeurs de conception du diamètre de garnissage, du type de matériau de garnissage et du rapport hauteur/diamètre du bioréacteur. Les résultats classés de la simulation ont été utilisés pour sélectionner les valeurs de conception optimales.

Contents

| | |
|--|-----|
| Synthesis..... | 17 |
| Analysis of a biomass torrefaction pilot plant | 26 |
| Predictive modeling of an anaerobic digester | 38 |
| Simulation of an upflow anaerobic filter | 51 |
| Study of upflow anaerobic filter treatment of municipal wastewater | 98 |
| A mobile data acquisition unit for environmental bioprocesses | 120 |
| Conclusions and perspectives..... | 124 |

Synthesis

Advances in data acquisition capacity, computing power and open-source libraries for data processing have opened a new and wide horizon for progress in the application of information systems knowledge and technology towards improving environmental quality. The range of possible environmental applications of IoT, big data, and machine learning includes studies of natural ecosystems, optimization of ecosystems services and the design of industrial machines to treat, for example, water, wastewater, municipal solid waste, and combustion and other waste gases. New approaches, skills and methods are needed to fully exploit this potential. My personal aims in undertaking a PhD were to change my perspective towards data acquisition and to learn how to use machine learning tools to extract more useful information from raw data. This thesis compiles three scientific publications that use machine learning methods to understand industrial processes that were designed to improve environmental quality.

In the **first article** titled “*Multivariate analysis of biomass torrefaction pilot plant operation data to support the development of a process model and controller*” I report the results of the analysis of data obtained during continuous torrefaction trials conducted in an integrated pilot plant (100 kg/h capacity). Torrefaction is a thermal process that yields solid, gas and condensable vapor phases. A process control strategy is required to specify and adapt the treatment conditions during a continuous process to assure process stability and to optimize the product yield and the energy efficiency of the process.

Why?

The purpose of the work was to extract information from the data that could be used to develop a process control strategy. An effective process control strategy would make it possible to accept raw materials having a wide range of physical and chemical characteristics and to process the raw material with minimal human intervention. In contrast, a process that requires a narrow range of raw material physical and chemical characteristics would require labor intensive presorting and a larger capacity for raw material storage. Consequently, the ability to accept raw materials having a wide range of physical and chemical characteristics is desirable because it would reduce investment and operating costs and increase the processing rate. Pilot plant data acquisition required many online and labor-intensive off-line measurements. An industrial scale plant might have even more on-line measurements. It might be possible to replace many off-line measurements by automatic measurements and machine learning based soft sensors. Development of an optimal process controller would require identification of the most important parameters influencing the process, the parameter values that indicate process instability and consequently need to be adjusted, and the lag-times between changes in manipulatable parameter values and the response. This knowledge would help to make the optimal selection of sampling points.

How?

The dataset acquired during pilot plant trials comprised 95 process variables including 23 on-line temperature measurements and 34 off-line measurements. The methods of Principal components analysis (PCA) and Partial least squares (PLS) regression were used to identify the most important parameters influencing the process. Parameter values that have a high positive correlation might be redundant. Thus, grouping these variables to create a new latent variable might be more informative than giving more importance to one member of the group, or than including all the members of the group as predictors in a model. The method of PCA approximates a complete data matrix by latent variables of lower dimension thereby removing redundant information. The method of PLS, sometimes called projection to latent structures, also makes use of the latent variable concept. The method of PLS regression maps a relationship between latent variables and a response variable. The regression coefficients are latent variables that group correlated predictors. The relative magnitude of the regression coefficients is an indication of the importance of the variables that are members of the group. In a stable process, the variation in parameter values about the expected values of the parameter is low. The Hotelling's t-squared statistic quantifies the total process variation at an instance. Plots of the Hotelling's t-squared statistics were used to identify the periods of process instability. The first principal component scores were then analyzed to identify the disturbance parameters and the manipulated parameters that contribute to the variance during these periods of process instability. The presence of significant and of persistent correlations between changes in the values of two parameters, was assessed by determining the correlations at different lag times. The parameters that had correlations with many other parameters were identified by calculating the total correlated lag times for each parameter pair.

Results

Principal components analysis of on-line operational data showed that the first 10 principal components account for approximately 80% of the total variance in the system. These are the most important indicators of process instability. Further analysis of the first principal component loadings identified the correlated parameters and the parameters that behave inversely during process operation. Analysis of the first principal component loadings identified the correlated parameters and the parameters that behave inversely during process operation.

A Partial Least Squares (PLS) regression model was developed to map the relation between the principal components obtained from process operational parameters and the Calorific Value of the torrefied product. The model does not explain 100% of the variation in the dataset because the number of the predictors used in the model is lower than the number of prediction vectors in the original dataset. One observes that adding more than 9 principal components does not increase the fraction of the total variance in the system that is explained by the model. It was found that the torrefaction temperature and the CO₂ content were important coefficients of the regression equation. Consequently, these parameters should be included in the process control strategy. In contrast, the remaining parameters showed less variation during the process operation. Consequently, some of these sensors and locations might be redundant and the measured values might have less impact on the process performance.

An average lag time of 41 and 34 minutes was found respectively between the thermal oil flow to the entry and exit sections of the torrefactor and many other process parameters. In contrast, the thermal oil flow to the central section of the torrefactor preceded changes in many other process parameters by an average of 24 minutes.

Conclusions

The results demonstrate the usefulness of the approach in understanding a complex system. For example, the results of this study demonstrate the importance of monitoring the torrefaction vapor CO₂ content and including these measurements in a process control strategy. The identification and the quantification of significant lag times between changes in the torrefaction thermal power supply and the product calorific value indicate that the mechanical design should be modified to reduce the response time and more precisely control the supply of thermal power to the process. The approach, methods and skills learned can be applied to other thermal treatment systems such as the pyrolysis of municipal solid wastes and combustion gas treatment.

In the **second article** titled “LSTM and 1-D Convolutional Neural Networks for predictive monitoring of the anaerobic digestion process” I report a comparison of 3 different artificial neural network model architectures trained on data acquired during 12 months of operation of an anaerobic digester treating food wastes. A novel control strategy based of the prediction of biogas production 4 hours in the future is proposed.

Why?

The purpose of the work was to make a predictive model to map process performance, evaluated in terms of CH₄ production, with practically measurable operational parameters. A predictive model could be used to automatically control the rate of waste food feeding to the digester. The automatic control strategy would help to intensify the anaerobic digestion process and to achieve the goals of small size, a short hydraulic retention time and long-duration autonomous operation.

How?

Principal Component Analysis (PCA), Independent Component Analysis (FastICA) and Partial Least Squares (PLS) regression methods were used for dimensionality reduction and to identify the parameters that make the greatest contribution to process stability. Three separate predictive model architectures were evaluated: 1) Long short-term memory (LSTM) 2) Convolutional (1-D CNN) and 3) a hybrid model featuring parallel data flow through LSTM and 1-D CNN layers. The process control strategy aimed to decide when to feed raw food waste to the digester. If the CH₄ production was predicted to increase, then the decision was to feed the digester.

Results

The daily feed volume, pH, ammonium concentration and volatile fatty acid concentration were identified as being the relevant parameters for use to train and test the Neural Network models. The 3 different model architectures were evaluated in terms of the RMSE between the predictions

and the true values. The LSTM and the 1-D CNN models proved to be accurate with RMSE values of, respectively, 0.017 and 0.026. However, model testing with new data showed that the LSTM did not have the ability to generalize. The hybrid model was not accurate (RMSE = 0.19). In terms of the frequency of correct feeding decisions the 1-D CNN model was found to have the highest rate of correct feeding decisions (58%) which was slightly lower than the rate obtained by skilled operators (63%).

Conclusions

The result of dimensionality reduction using PCA, ICA and PLS methods was the selection of only 3 predictors for use in the ANN model. This is practically useful for the selection of the relevant process sensors. However, the observed low accuracy of the decisions to feed suggests that the data acquisition strategy should be modified to include additional process sensors or a higher sampling rate. Alternatively, the ANN model could be improved to make better use of the data. Although the predictive model built using 1-D CNN architecture was slightly less accurate than the LSTM model, the state image approach is intuitively interesting and merits further development for use in process control. Moreover, the 1-D CNN model proved to more accurately predict future biogas production. Considering that the acquired training data is the result of the decisions made by expert operators, the capacity of the model to predict future conditions should be compared to the capacity of the skilled operators to predict future conditions. The comparison is, thus, the ratio of correct decisions made using the model to correct decisions made by the operators. This ratio is 58%/63% (0.92). In conclusion, the control strategy using the 1-D CNN model was nearly as accurate as feeding control by skilled operators.

In the **third article titled** “An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process” I built an artificial neural network predictive model based on data acquired during 7 months of continuous operation under industrial conditions and on supplementary mock data. I used the model as a simulation tool to assess different design values of the anaerobic filter packing material diameter, packing material type and the bioreactor height to diameter ratio. The study hypothesis was that different sets of design values can be distinguished by simulation of previously untested data. Thus, the design values of the industrial equipment would be selected based on significant differences in the mean response values of simulation runs. Additionally, I proposed to build a mobile data acquisition system and use an experimental approach that aims to save time and reduce costs.

Why

The conventional methods of modeling wastewater treatment processes involve the use of controlled experiments under strictly defined laboratory conditions and the use of established mechanistic models. For example, the ADM-1 model¹ is implemented as a differential and algebraic equation comprising 26 component dynamic concentration variables and 8 implicit algebraic variables that describe a system of subprocesses most of which are microbe mediated biochemical transformations defined by stoichiometry and kinetic coefficients. Quantification of mass flows is

simplified by lumping components into classes such as proteins, Chemical Oxygen Demand (COD), which is a practically useful expression of pollution strength as the mass of oxygen required to oxidize a pollutant, soluble carbohydrates, Long Chain Fatty Acids (LCFA), etc. Use of the model involves the determination of the kinetic coefficients and chemical species by time and resource intensive laboratory studies. In practice, to save money and time, many of the required values are selected from the literature.

Although the mechanistic models and the associated research to develop them have served to improve the understanding of the system of biochemical processes involved in anaerobic digestion, the models do not account for important features such as the bioreactor mechanical design, the hydrodynamic parameters and the large and frequent variations in the influent characteristics, microbial populations and operating conditions that are often encountered in industrial conditions. Based on my own professional experience, I identified several inadequacies in this approach and then aimed to make an improved approach to the design of environmental bioprocesses.

For example, the conventional approach might not adequately account for the variation encountered under real industrial, and especially under real environmental, conditions where it is not possible to continuously measure and regulate the substrate composition and mass flow. I imagined that it would be more efficient to build a predictive model of a swarm of data points collected under real industrial conditions.

The standard models require the determination of biochemical parameters such as kinetic constants, inhibition factors and chemical intermediates. Acquisition of this data requires laboratory experimentation under artificial conditions. In practice many of the required values are assumed and the effects of actual industrial operating conditions are not studied. The conventional workflow to develop environmental bioprocesses could be improved, for example, by reducing the time and resource requirements for laboratory experimentation and by collecting data under real operating conditions. Additionally, data collection systems could be designed to make it possible to easily vary and measure real mechanical parameters under real operating conditions.

In the case of the upflow anaerobic filter (UAF) studied, the goal was to determine the design values of the packing material spherical diameter, type of packing material and fixed bed height to diameter ratio must be specified. Finding the optimal combination of design values using conventional methods would be time consuming and costly.

How

The use of Artificial Neural Networks is a black-box approach to modeling complex systems. Using a black-box approach made it possible to not include the conventional bioprocesses parameters such as kinetic constants, microbial inhibition factors, and hydrodynamic parameters that are difficult to measure. Instead, easy to measure parameters like influent flow rate, exit gas composition and the dimensions of constructed mechanical features were the only predictors used in the model. Following recent developments in image recognition methods, the ANN model was highly overparameterized to make it possible to simulate by interpolation using new predictors. The ratio of model parameters to data was 35.

Data were acquired using a test unit installed in a municipal wastewater treatment plant and continuously fed primary effluent. The experimental dataset was acquired before deciding to use an artificial neural network modeling approach. The predictors were a small dataset of 3 mechanical parameters. Since ANN models require large datasets, I created surrogate data and I created mock experiments to supplement the experimental dataset. The use of mock data was necessary for this proof-of-concept study. However, mock data should not be used to design a commercial bioreactor. The proposed method to improve the design process of a commercial bioreactor would require a dataset comprised of actual measurements. Next, I made numeric models and compared the prediction accuracy of a 4th degree polynomial, a multi-layer perceptron (MLP), and an LSTM model. The MLP model proved to be the most accurate.

The MLP model was saved and reloaded for use as a simulation tool to aid in designing the UAF bioreactor. The use of simulations during the design process is a common practice to test different designs. The final design specification can be determined after comparing the simulated test cases. A full factorial design was used to specify 125 permutations of the 3 mechanical parameters (factors) at 5 levels. To make comparison between simulated experiments possible, the same daily influent flow rate profile was used to run all 125 test cases. I ran 20 replicates of each simulated experiment. The experiments were then ranked according to average total calorific value (CV) reduction.

Results

The coefficients of determination (R^2) of the MLP and of the 4th degree polynomial models were, respectively, 0.66 and 0.37. The prediction accuracy of the MLP, evaluated in terms RMSE between predicted and observed calorific value reduction during validation tests, was 10.1%. I built, trained, and evaluated all possible MLP architectures having between 1 to 13 hidden layers and from 8 to 4096 units per hidden layer. I observed that the models having between 64 and 2048 units per layer had the highest accuracy and that using between 2 and 6 hidden layers yields similar results. The results of 125 simulations were ranked by total CV reduction. The ranked CV reductions covered a wide range with the highest ranked value more than 3 times greater than the lowest ranked value. The very low *p-value* of the difference between the means of the highest ranked CV reductions made it possible to select the design values for the industrial UAF bioreactor. This result demonstrates that a dataset generated during a 6-month experimental phase can be used to make simulations covering more than 60 years of continuous operation.

Conclusions

Machine learning and artificial neural networks are already well-established methods to model environmental bioprocess data. The originality of this work lies in the use of a neural network for simulation of a designed bioprocess. The paper described, for the first time, a new workflow and the use of a MLP as a simulation tool to specify the constructive features of an Upflow Anaerobic Filter. Unlike conventional experimental designs, the proposed workflow allows the physical-chemical parameters, such as stream composition, and the temporal factors to vary within their real ranges. Thus, the use of Artificial Neural Network predictive models for simulation simplifies

experimental planning and data acquisition because the requirement is only to build a wide ranging and dense dataset that covers the actual industrial conditions in the field. The design values of mechanical predictors were set according to an experimental plan based on the orthogonal Taguchi arrays. The acquired and mock data were used to build a highly parameterized neural network model. The study hypothesis that different sets of design values can be distinguished was tested by comparing the mean performance attained by each set of 125 different simulated experimental conditions. The conclusion is that the set of mechanical predictor values that resulted in the highest mean performance are the best design values for the anaerobic digester.

The use of artificial neural networks is intended to complement existing DEA models and other mechanistic approaches to modeling. The reliability or trustworthiness of a neural network model is assessed by evaluating the error between the predictions and the measured data. This method is a powerful way to validate the predictive capacity of the model and the conclusions obtained using the model. However, artificial neural networks are black-box models that do not contribute to our understanding of the underlying process physics and chemistry. Consequently, our ignorance about the real biochemistry is still a limitation of artificial neural network predictive models. In contrast, the mechanistic models have the merit of trying to be complete representations of the processes.

Motivation for writing this thesis

The motivation for writing this thesis comes from my personal dissatisfaction with conventional modeling approaches, the observation that Artificial Neural Networks have demonstrated their capacity and usefulness as a prediction tool to solve non-linear and complex problems in engineering applicationsⁱⁱ, recent increases in computing power, and the availability of open-source libraries that made it possible for me to build models to map complex, non-linear relationships between input and output data.

In all three articles I used conventional machine learning methods to extract latent information from datasets and build predictive models. The motivation for writing the first article came from my personal discovery that the conventional machine learning toolbox contains powerful open-source methods and libraries that could be practically useful for analyzing the dataset obtained from torrefaction pilot plant operation. Similarly, the motivation for writing the second article came from my personal discovery of the wide variety of possible ANN architectures and the intuition that the CNN state-image approach was suited to modeling the anaerobic digestion process. The motivation for writing the third article came from my intuition that a dense dataset acquired under nearly industrial conditions could be used to make an ANN that would be useful for simulation. To test this hypothesis, a MLP predictive model of an upflow anaerobic filter was built and used for simulation. Thus, the innovative content and contribution of this thesis is: 1) a demonstration of a new approach that does not require designed laboratory experiments, 2) an additional application of machine learning tools to environmental processes, and 3) the use of an MLP model for simulation and selection of mechanical constructive features.

Discussion of artificial neural network versus mechanistic approaches to modeling environmental bioprocesses

Since an artificial neural network is always a “black box”, the user renounces to obtaining knowledge about the physical, chemical, and biological mechanisms of the modeled process. In contrast, the mechanistic approach aims to describe, in precise terms, the physical, chemical, and biological mechanisms of the modeled process. In my opinion, these two approaches are complementary. Work to develop and improve both ANN and mechanistic models should continue. When used with machine learning methods, the ANN approach identifies and uses the most relevant predictors to obtain a robust model that is often more accurate than a mechanistic model over a wide range of operating conditions. The accuracy is due to using the basic machine learning method of training the model using a very large dataset, calculating the error between predictions and true values, and then adjusting the model parameters to improve the accuracy of the predictions. The most relevant predictors are not necessarily the same as those used in a mechanistic model. Moreover, the most relevant predictors might be unique characteristics of the equipment design and not reveal anything about the underlying physics, chemistry, or the biological processes. Consequently, one can build an accurate predictive model of an environmental process without knowing or learning anything about how the process works at the physical, chemical, or biological level. Thus, the ANN approach is often practically useful for engineering applications without contributing to theoretical understanding.

The mechanistic modeling approach, however, attempts to build a system of interrelated equations each describing precisely defined physical, chemical, or biological sub-processes. The development of this approach involves scientific discovery and theory. The strength of this approach lies in the theoretical understanding of the subprocesses and their interrelationships. This understanding of basic mechanisms might be useful in the design new or optimized processes because the designers will know where to focus their attention and experimental investigations.

The ANN method is data driven. Usually, the more data available, the more accurate the model. Additionally, the model can continuously improve with the use of new data. In contrast, the mechanistic model is built using experimentally acquired data and has a more limited capacity to integrate new data acquired during industrial operations. Given the current trend of increasing operational data acquisition, the power of ANN and machine learning approaches will certainly increase.

Wider application of the results

The machine learning methods that I learned while preparing this thesis can be applied to any complex system. For example, PCA and PLS are practical methods to gain insights into complex datasets, to detect deviations from normal conditions, to identify the most important predictors, and for root cause analysis. This thesis focused on machines. However, the methods can be applied to improve designed natural systems such as forests, wetlands, lakes, and oceans. Natural ecosystems can be modified and controlled to provide economically valuable ecosystem services that promote, for example, water and air purification, fish populations, pest control, and resistance to fires.

Availability of computer code and data

The Matlab® and Python code used to analyze the data and make the models is available on request. The data that support the findings of the torrefaction and the anaerobic digestion pilot studies are available pending approval of the owners. The data that support the findings of the upflow anaerobic filter study are available from the author on request.

ⁱ Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, Sanders WT, Siegrist H, Vavilin VA. The IWA Anaerobic Digestion Model No 1 (ADM1). *Water Sci Technol.* 2002;45(10):65-73.

ⁱⁱ Yilmazkaya E., Dagdelenler G., Ozcelic Y., Sonmez H. Prediction of mono-wire cutting machine performance parameters using artificial neural network and regression models. *Engineering Geology* 239 (2018) 96–108.

MULTIVARIATE ANALYSIS OF BIOMASS TORREFACTION PILOT PLANT OPERATION DATA TO SUPPORT THE DEVELOPMENT OF A PROCESS MODEL AND CONTROLLER

Mark M. McCormick^{1*}, Boris Correa²

¹ University of Applied Sciences Western Switzerland, HEIG-VD, Industrial Bioenergy Systems Unit, Avenue des Sports, 20, CH-1401 Yverdon-les-Bains, Switzerland, e-mail: mark.mccormick@heig-vd.ch

² GRT Technologies and Engineering (GRT) SA, Rue des Ducats 40 B, 1350 Orbe, Switzerland

Keywords: Torrefaction, Biomass energy, PCA, PLS regression, Process control

Abstract. Torrefaction of biomass is a thermal treatment process executed in an oxygen poor environment in order to improve storage and combustion properties. This paper reports the results of multivariate analysis of data obtained from a continuous torrefaction trial conducted in an integrated pilot plant (100 kg/h capacity) comprising automated feeding, drying, torrefaction, torrefaction gas combustion and product cooling. The purpose of the work was to gain insights into the process behaviour that might be useful in the conception of a process controller that makes it possible to obtain a product having a constant calorific value from heterogeneous raw biomass having a fluctuating calorific value. The questions addressed in this paper are; what are the most important predictors of final product calorific value, what parameters should be manipulated, what is the lag time between parameter adjustment and response, which manipulated parameters account for deviations from mean process values during operation and what is the predicted product calorific value as a function of the important manipulated parameters? Time series data collected during 35 hours of continuous operation and composed of observations of 95 process variables including 23 on-line temperature measurements and 34 off-line measurements of raw and torrefied biomass properties such as moisture content and calorific value were analysed. The results of PCA and PLS analysis identified the dryer temperature as having the most important influence on process stability and the temperature of the central torrefactor section as having the most important influence on product calorific value.

1 INTRODUCTION

In the global effort to reduce fossil CO₂ emissions and to mitigate climate change, biomass is considered to be an important renewable raw material for the production of fuels and chemicals. However, the use of biomass is complicated by its instability and inhomogeneous proximate and ultimate parameters such as moisture, volatiles and oxygen content. Torrefaction has been proposed as an effective solution to problems such as biomass instability during storage, high grinding energy requirements and inefficient combustion and gasification due to inhomogeneity of the fuel. To study and address these challenges, a complete pilot biomass torrefaction system was built and used to tests a variety of feedstocks including hard wood chips, forest residues and branches from urban tree maintenance.

An overview of the values measured during the pilot torrefaction trials was obtained from the plots of time-series sensor data. However, given the large number of on-line and off-line measurements and the differences in the types of measured properties, it is difficult to make practically useful conclusions about the process control strategy from these observations. Alternative techniques of methodical analysis of on-line and off-line data are required to identify the relevant measured and manipulated variables and to rank them in terms of their explanation of process variance and correlation thereby leading to useful insight into the process and contributing to the development of strategies to control the process.

As a complement to feed-back control which compensates for predictable disturbances, multivariate analysis of process data is used to detect and diagnose occurrences of special events and then to remove them by improvements to the automatic control system [1].

This paper reports the first PCA and PLS analysis of the pilot plant operation data. Principal components analysis (PCA) approximates a complete data matrix by latent variables that are matrix products of lower dimension thereby removing redundant information. This objective is realized by creating lines (principal components) and planes that are least squares approximations of the data points that make the variance of the coordinates on the line or plane as large as possible [2]. In a similar manner, Partial Least Squares (PLS) regression also known as Projection to Latent Structures (PLS) establishes a relationship between the predictor variables and the response variable(s) [2]. PCA and PLS are useful methods to obtain an overview of data and the relationships between blocks of data in the early stages of a project [2]. In an industrial process, tens of thousands of data points are easily collected and all of these data should be used to effectively monitor and diagnose process performance. However, only a few underlying events are driving a process at any given time. By examining process data from a large number of measured variables in the projection space defined by a small number of latent variables it is often possible to extract useful information that can be used to improve the process [1]. PLS has the ability to analyze data with many, noisy, collinear, and even incomplete variables [3]. Another useful procedure to monitor a multivariate process consists in comparing the Hotelling T^2 statistic obtained from principal component scores to a value of T^2 corresponding to a confidence interval obtained from an F-distribution [4]. Lestander [5] studied memory effects in wood pelleting plant by evaluating the cross correlations between process operational parameters and off-line measurements of product qualities such as pellet dryness. Calculating the cross-correlations resulting from simulated time delays between process parameters gives an indication of lag-times between changes to manipulated parameters within the process and their effect on other process parameters.

The methods and results described in this paper can be used to achieve the process control goals of increasing the product calorific value and production rate while at the same time widening the acceptance limits of the biomass raw material quality.

2 PILOT PLANT AND TRIALS

2.1 Pilot plant description

The pilot plant was designed to accept and thermally treat shredded or chipped woody biomass having a moisture content of up to 35%. Biomass is continuously transferred by belt conveyor from a 10 m³ feed hopper to a counter current tumbling dryer that is heated by exhaust gas mixed with air and by a thermal oil jacket. Dried chips are transferred via a rotary valve to a screw-conveyor torrefactor that is heated by thermal oil. The torrefied product is cooled by a water mist, then further cooled during transfer through a water-jacketed cooling screw to a hopper scale, and finally discharged into 1m³ big-bags. Torrefaction gas flows through a heated pipe to a Flox® burner system with a turbo-compressor. Natural gas is blended with torrefaction gas to compensate for the lack of thermal energy supplied from torrefaction gas combustion. The plant is equipped with an HMI, PLC and data logger designed to provide completely automated continuous operation except for manual loading of the feed hopper and handling of the big bags.

2.2 Process sensors and on-line measurement

All of the measured variables were logged at 5-second intervals. Temperature sensors are either type K thermocouples or Pt100 temperature detectors (Endress-Hausser, CH). An IR sensor (Raytek, USA) measures the wood chip surface temperature at the dryer exit. The CO₂ and O₂ content of the torrefaction gas were measured using respectively an IR sensor (Dynamet, UK) and a lambda (Bosch, DE) probe. The flow rate of hot gas entering the dryer was measured using a Pitot tube (Kimo, F).

2.3 Off-line measurements

Raw materials and torrefied product were characterized according to ISO specifications for solid biofuels [6]. Water consumption, cooling water temperatures and flow rates, thermal oil flow rate and the ambient temperature were measured or calculated off-line. Natural gas flow rate was confirmed by an off-line volumetric gas counter.

2.4 Process control

The primary objective of the process controller is to assure that fresh biomass is fed at the rate required to produce torrefied biomass at the requested rate. The strategy to attain this objective involves control loops that link system sub-processes that are controlled individually by PID process controllers. The dryer PID controller aims to achieve a temperature set point by controlling methane flow to the burner, the thermal oil and hot gas flow rates to the dryer and the torrefaction residence time. The torrefactor PID aims to attain a temperature set point by controlling thermal oil flow to the torrefactor. The control of auxiliary systems and safety features are not described in this paper.

2.5 Torrefaction trial procedure

Biomass raw materials were sampled at the time of delivery and again from the point of belt conveyor loading at between 1 and 2 hour intervals during pilot plant operation and then characterized after the torrefaction trial. Given the long heat-up and cool down times, 20 m³ of raw biomass is the minimum amount tested during a trial. Natural gas only is supplied to the burner for ignition and during the following 4 hours that are required to bring the pilot plant to the operating temperature of approximately 250° in the torrefactor after which natural gas and torrefaction gas are co-combusted. The initial biomass loading rate is set to less than 50% of the capacity and then gradually increased. Process set points and PID controller values were set manually based on previous experience and on observations made during the current trial. Grab samples of the product were manually obtained at intervals of between 1 and 2 hours.

3 DATA AND ANALYSIS

3.1 On-line data

Since the on-line data have units of °C, amperes, Nm³, kg, kg/h, bar, mbar and %, the data were unit-variance scaled by dividing each column by its standard deviation thereby assigning equal weights to each data column. The data were mean centered to move the center of the data to the origin of the variable coordinate system. Unit-variance scaling combined with mean centering is called auto-scaling [2]. Since torrefaction is intended to be a continuous process, only the data acquired between the time of the first appearance of torrefied product and the beginning of cool down were considered in this study. Table 1 lists the on-line probes considered in this study and their corresponding short names and i.d numbers.

| Probe ID | Parameter | Param. N° | Probe ID | Parameter | Param. N° |
|----------|---|-----------|----------|--|-----------|
| F101PV | Biomass feed rate | 1 | T403bPV | Final torrefactor heating segment | 21 |
| T201PV | Initial dryer heating segment | 2 | T403sPV | Torrefactor final chamber | 22 |
| T202PV | Middle dryer heating segment | 3 | T404PV | Thermal oil temperature, torrefactor | 23 |
| T203PV | Final dryer heating segment | 4 | T405PV | Immersed in wood chips, torrefactor exit | 24 |
| T203sPV | Final dryer heating segment | 5 | T406PV | Torrefaction gas, burner entrance | 25 |
| T204PV | Dryer thermal oil | 6 | T407PV | Exhaust gas, burner exit | 26 |
| T205PV | Post dryer exit gas | 7 | T408PV | Exhaust gas, between the heat exchangers | 27 |
| T206PV | Heat exchanger exit gas | 8 | P401PV | Torrefaction gas pressure | 28 |
| T207aPV | Wood surface, dryer exit | 9 | P402PV | Thermal oil reservoir pressure, torrefactor | 29 |
| T207bPV | Immersed in wood chips, dryer exit | 10 | Q401PV | O ₂ , torrefaction gas | 30 |
| T208PV | Hot gas entering dryer | 11 | Q402PV | CO ₂ , torrefaction gas | 31 |
| F202PV | Gas flow rate entering dryer | 12 | TO401IPV | Current, torrefactor motor | 32 |
| SE201FPV | Current, dryer motor | 13 | VR401PV | Thermal oil by-pass valve, torrefactor | 33 |
| VT201IPV | Current, rotary valve between the dryer and torrefactor | 14 | VR402SP | Thermal oil by-pass valve, torrefactor central section | 34 |
| VR201PV | Thermal oil by-pass valve, dryer | 15 | VR403SP | Thermal oil by-pass valve, torrefactor final section | 35 |
| VR202SP | Thermal oil by-pass valve, dryer central section | 16 | T501PV | Interior cooling screw, midway | 36 |
| VR203SP | Thermal oil by-pass valve, dryer final section | 17 | T502PV | Interior cooling screw, exit | 37 |
| T401PV | Initial torrefactor heating segment | 18 | T601PV | Turbo exit | 38 |
| T402PV | Middle torrefactor heating segment | 19 | F601PV | Methane flow rate | 39 |
| T403aPV | Final torrefactor heating segment | 20 | VE601IPV | Current, dryer ventilator motor | 40 |

Table 1. on-line parameters

3.2 Off-line data

The quality of the raw material and the torrefied were determined after the torrefaction trial. Table 2 summarizes the main characteristics of the biomass raw material and the torrefied product.

| Biomass quality parameter | Unit | Raw material | | | Product | | |
|------------------------------------|-------|--------------|-----|--------|---------|-----|--------|
| | | Mean | Std | CV (%) | Mean | Std | CV (%) |
| Humidity | % | 24.0 | 1.5 | 6.2 | 5.8 | 3.2 | 55.0 |
| Volatiles content | % | 81.7 | 0.3 | 0.3 | 77.3 | 2.3 | 2.9 |
| Ash content | % | 1.0 | 0.2 | 18.6 | 0.9 | 0.2 | 25.0 |
| Fixed carbon | % | 17.3 | 0.1 | 0.6 | 21.8 | 2.1 | 9.7 |
| Upper Calorific Value, dry | MJ/kg | 19.9 | 0.2 | 0.8 | 21.0 | 0.4 | 2.1 |
| Lower Calorific Value, as received | MJ/kg | 13.6 | 0.4 | 3.1 | 18.6 | 0.9 | 4.8 |
| Lower Calorific Value, dry | MJ/kg | 18.7 | 0.1 | 0.8 | 19.9 | 0.5 | 2.3 |

Table 2. Biomass properties

3.3 Principal Component Analysis of on-line data

Figure 1 shows the percent of the total variance in the on-line measurements explained by each principal component. The first principal component accounts for only 25% of the variance in the system and the first 10 account for only 75% of the total variance. This result indicates that the pilot system has many non-correlated measurements and few redundant measurements. Figure 2 shows that 4 principal components explain 75% of the variation in the temperature data. This indicates that some of the temperature probes might be redundant.

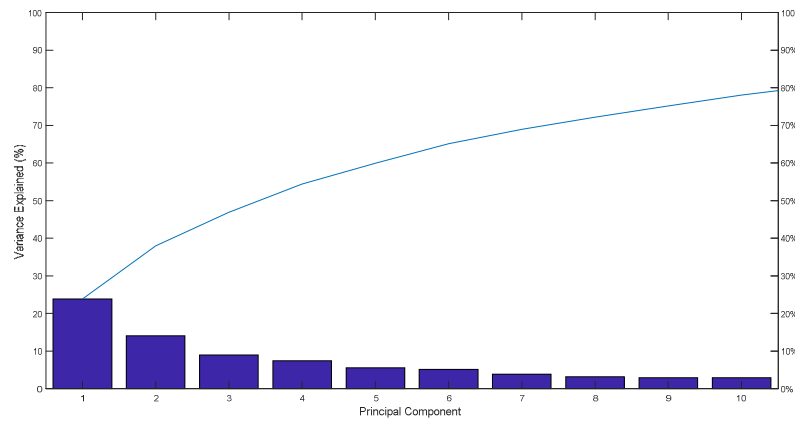


Figure 1. Variance in all on-line data explained by principal components

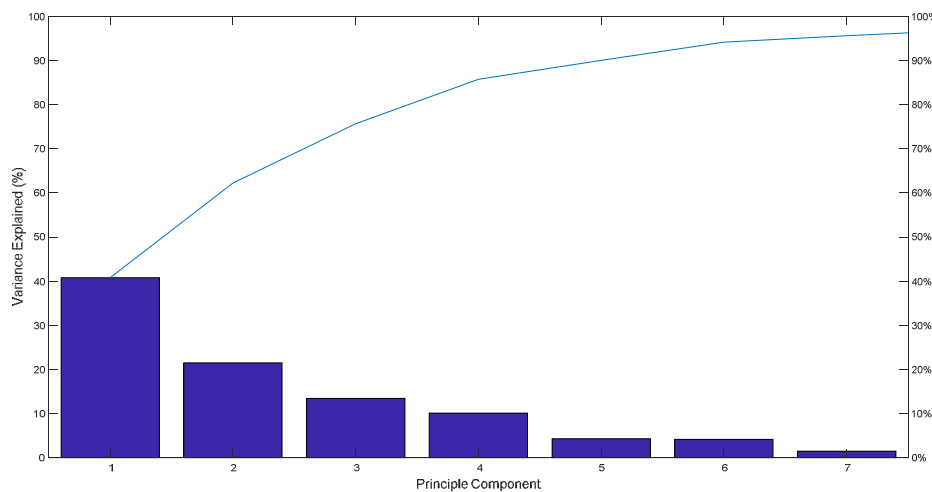


Figure 2. Variance in temperature data explained by principal component

3.4 Process instability

The Hotelling T-squared statistic is a measure of the multivariate distance of each observation from the center of the data set [7]. A time series plot of the calculated T-squared statistic and the alpha value calculated from the F-distribution reveals process instability as the periods of large deviations from the center of the data set when the value of the T-squared statistic is greater than the value of alpha. Score plots produced at the times when T-squared exceeds alpha identify the particular sensors and the manipulated parameters that contribute to the variance during periods of process instability.

3.5 PLS analysis of product calorific value

PLS is often used to predict the values of time-consuming off-line physical sample measurements in real-time using the online data from the process. PLS is a latent variable method having the objectives of providing the best explanations of the predictor and response variable spaces and in predicting the greatest relationship between the predictor and response variable spaces [2]. In this study the predictor space is comprised of the 40 on-line parameters listed in table 1 and the response space is the lower calorific value of the dry torrefied product.

3.6 Correlation analysis of dryer heating

Since the results of principal component analysis of all of the process data showed that the dryer contributes the greatest amount of overall process variance, the dryer was further analyzed by determining the correlations between dryer operating parameters including the biomass feed rate.

3.7 Lag times between on-line parameters

Lag times between all possible combinations of 40 on-line operational parameters were determined using the Matlab® [7] *finddelay* function embedded in for loops. The function creates time delays between parameters, evaluates the resulting cross correlations and returns the delay in terms of the difference in the sample row numbers that results in a significant cross correlation.

4 RESULTS AND DISCUSSION

4.1 Principal Components Analysis of on-line data

Analysis of the first Principal Component loadings show that the variance in final dryer section (T203s) contributes the most to overall process variance. Taken together, the dryer temperature probes make a greater contribution to process variance than the torrefactor temperature probes. However, the difference is not great and the difference might also be explained by the position of the probes in the dryer and the torrefactor.

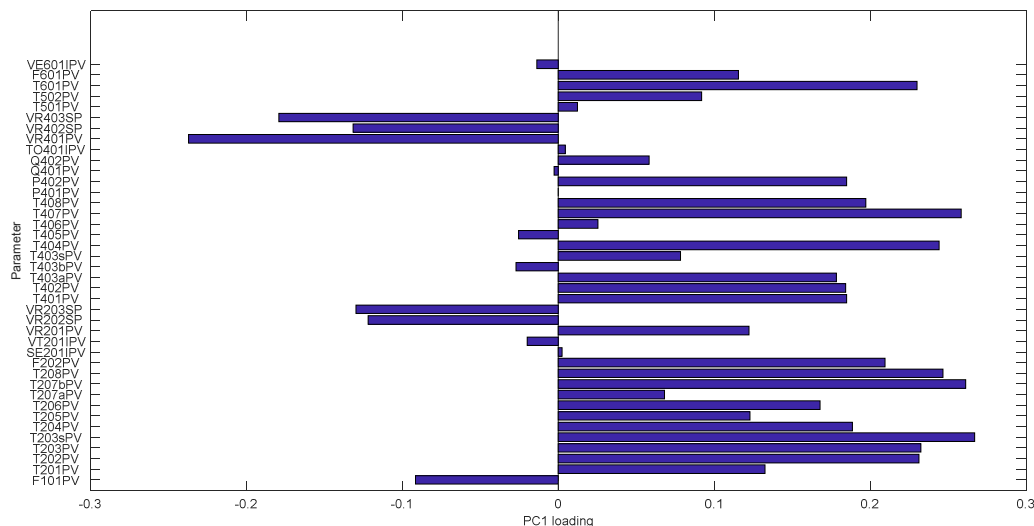


Figure 3. First Principal Component loading

Figure 4 plots the first and second principal component loadings. The parameters that have loadings located in diagonally opposite quadrants behave inversely during process operation [2]. When the value of one measured parameter increases, the value of the parameter in the diagonally opposite quadrant decreases and the greater the separation on the plot, the greater the opposing effect. The greatest number of scores are located in the right

quadrants indicating that many measured values are positively correlated. The location of the measured parameters numbers 1, 16, 17, 33, 34 and 35 in the lower left quadrant indicates that when the measured values increase, the measured values of the other parameters decreases. Measured parameter number 1 is the biomass feed rate and numbers 16, 17, 33, 34 and 35 are the fraction of opening of the valves controlling thermal oil flow into the dryer and the torrefactor. One can conclude that when the raw material feed rate increases, the thermal oil valves open and the temperatures throughout the system decrease.

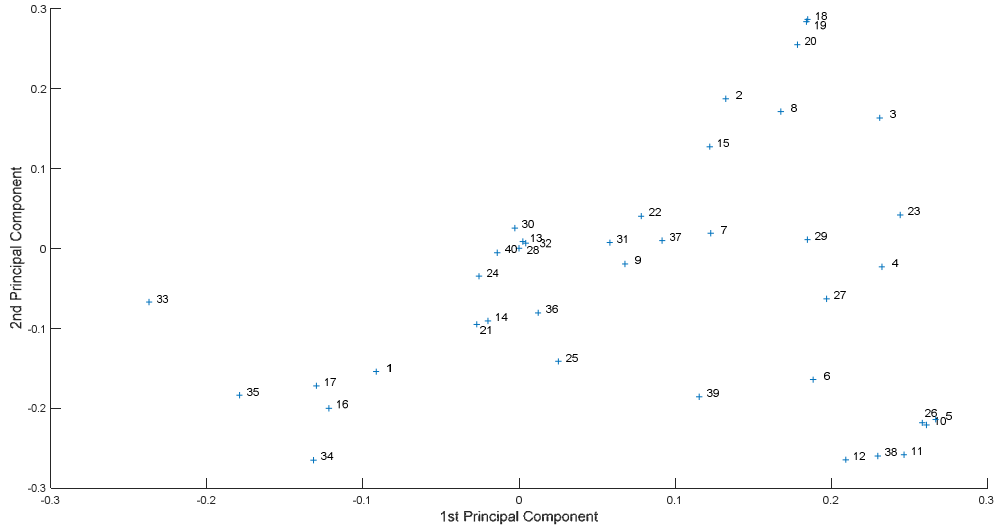


Figure 4. Loading plot, Principal components 1 and 2

4.2 Process instability

Periods of process instability can be defined as periods when the T-squared statistic exceeds an alpha limit of 5%. Figure 5 shows the T-squared statistic calculated from the first principal component scores of 40 on-line parameters.

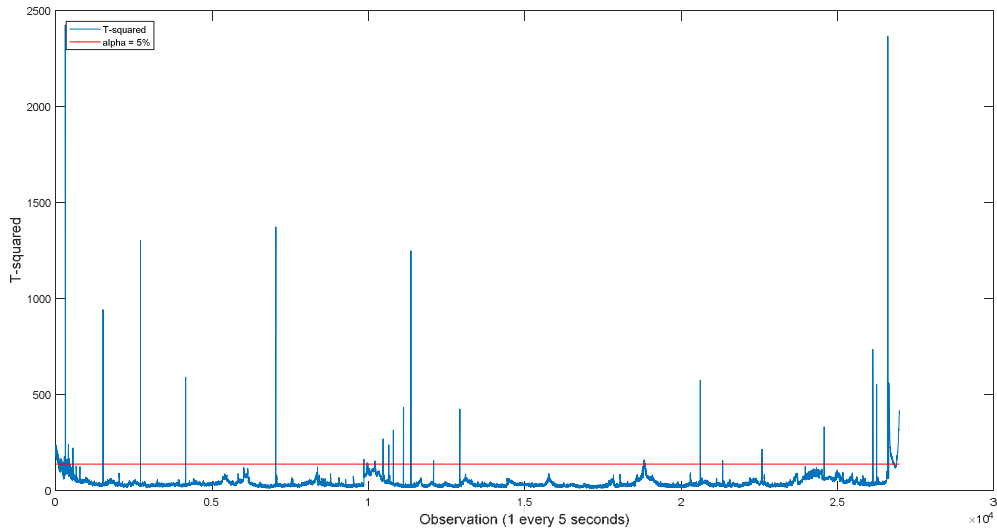


Figure 5. T-squared

The value of the T-squared statistic increases around observations 10'000 and 18'000 indicating high variance of the on-line measurements and process instability. In order to identify the parameters that deviate the most from the central values during process instability, score plots of the measured parameters at observations 10'025 and 18'700 were examined. The score plots show that the highest scores during both periods of instability were those of the dryer temperatures, in particular the temperature of the final dryer section just before transfer to the torrefactor.

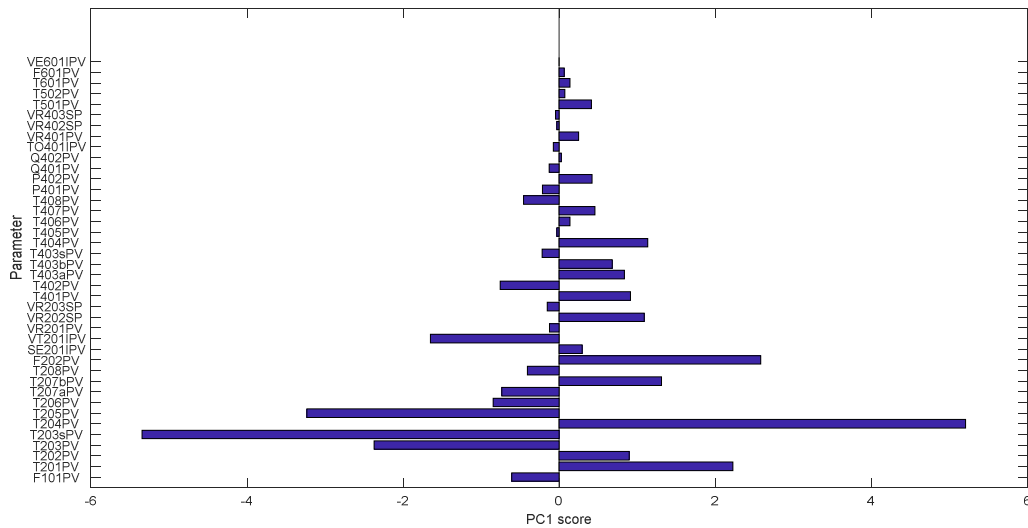


Figure 6. First Principal Component scores, observation n°10'025

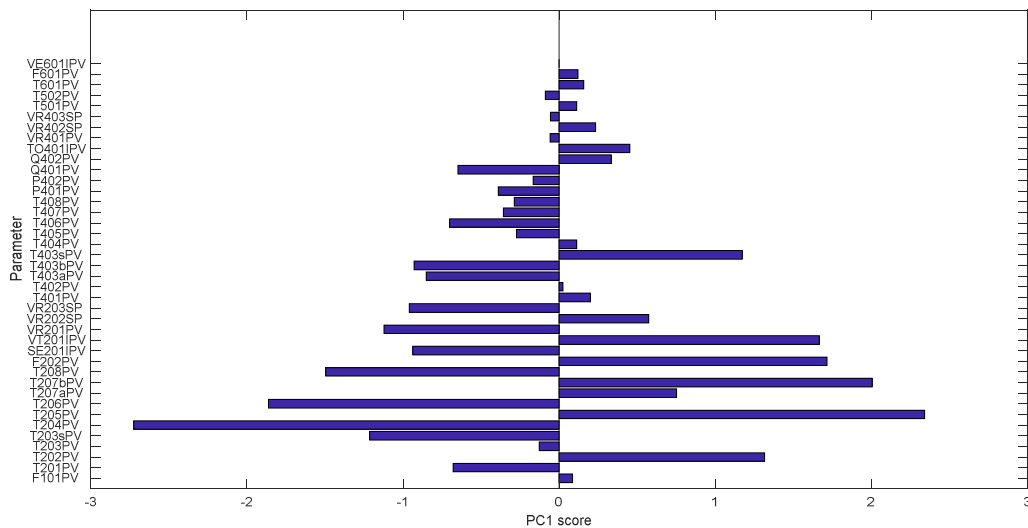


Figure 7. First Principal Component scores, observation n°18'700

4.3 Predictors of product quality

Provided that the product respects the limit values for environmental contaminants, the Net Calorific Value (NCV) is the most important quality indicator of the of the torrefied product because the NCV indicates the amount of thermal energy produced in a typical non-condensing biomass burner and is influenced by the moisture content of the fuel. According to ISO 17225:8, the NCV must always be greater than 21 MJ/kg. The Gross Calorific Value (GCV) includes the latent heat of condensation of the combustion products and indicates the chemical energy stored in the fuel independently of the fuel moisture content. Since variations in product humidity during the pilot trials is largely explained by the spray cooling system and improvements should reduce the variation in product moisture content, the product quality from the pilot plant was evaluated in terms of GCV. A linear expression involving beta values for each process parameter was obtained by PLS regression of 40 on-line process parameters and the product GCV. R-squared is 0.76.

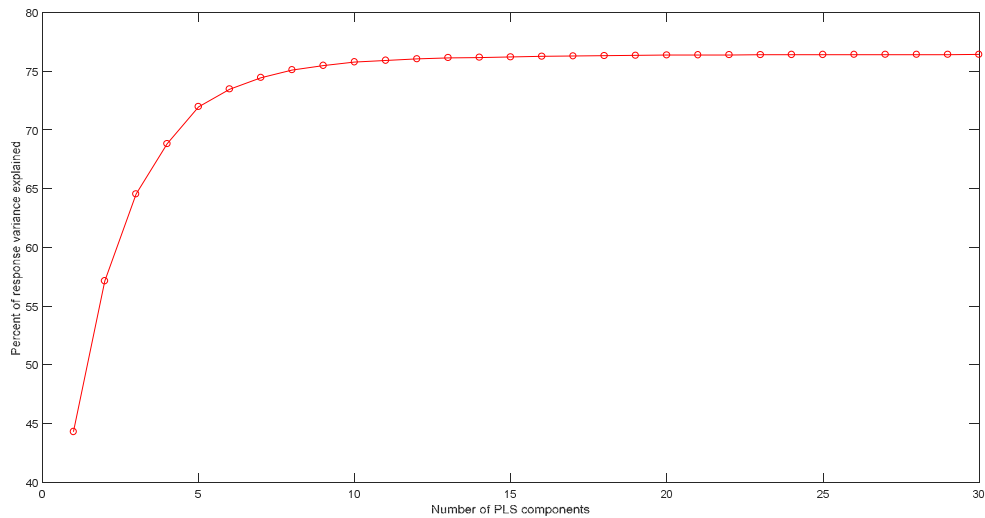


Figure 8. Response variance explained by PLS predictor components

The beta-values shown in figure 9 show the relative importance of each predictor parameter in determining the product GCV.

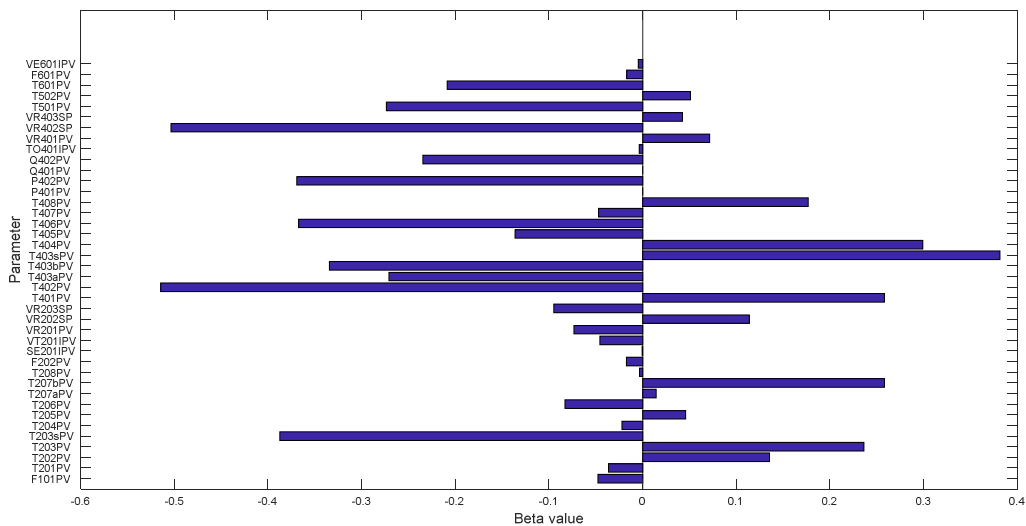


Figure 9. PLS β values, Product UCV as a function of on-line process parameters

As indicated by the relative values of β , the torrefactor temperature values have a more important influence on GCV than the dryer temperature values. Interestingly, the biomass feed rate and the methane feed rate do not have an important influence on the product GCV while the CO_2 concentration of the torrefaction gas has an important influence on product GCV. PLS regression of on-line data and the off line determinations of the calorific value of the torrefied product were used to make a predictive model.

$$\text{UCV, dry} = 1.7 + X \cdot \beta + Y_{\text{residual}} \quad (1)$$

Where X is the matrix of on-line measurements, β is a vector of coefficients and Y_{residual} is a vector of residuals returned by the Matlab[®] plsregress function. The Y-intercept of 1.7 was also returned by the Matlab[®] plsregress function. The model was cross-validated with a separate torrefaction batch by comparing the values predicted from the on line measurements to the measured calorific value of the product (Figure 10). Using the first 9 principal components results in the model having the best fit to the off-line data.

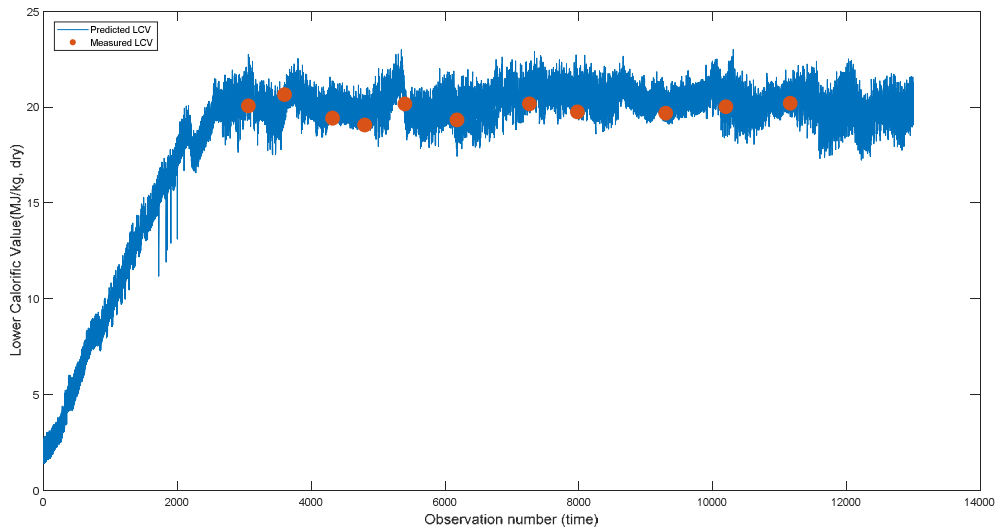


Figure 10. Cross validation of the PLS regression model

4.4 Dryer heating

Since biomass drying requires by far the greatest amount of thermal energy consumed by the process, the correlations between manipulated parameters and dryer outlet temperature were examined in order to identify ways to optimize the dryer and the control of the dryer. Figure 11 shows the correlations between the dryer operating parameters including the biomass feed rate. The code numbers are explained in table 1. The temperature measurements of probe 207b are highly correlated with the temperature and the flow rate of the hot gas entering the dryer. The biomass feed rate is inversely correlated with the temperature in the middle of the dryer suggesting that the feed rate is limited by the temperature of the central section of the dryer.

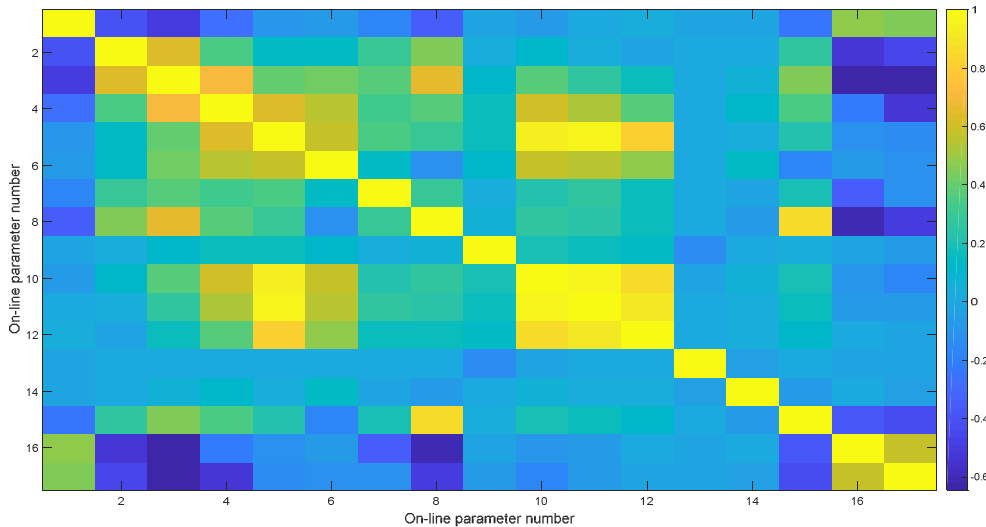


Figure 11 Correlations between dryer operating parameters

4.5 Lag times between sub-process

Figure 12 shows the 40 on-line measured parameters on the horizontal axis. The vertical axis shows all of the parameters having significant lags for each of the measured parameters. The height of each bar corresponds to the lag duration. The existence of a lag implies a significant and persistent correlation between changes in the values of two parameters during the torrefaction run. Positive values indicate that the parameter lags the parameters shown on the horizontal axis. Negative values indicate that the parameter precedes the parameters shown on the horizontal axis. In other words, the parameter is regularly changed after observing certain values of other measured parameters. Knowing the duration of the lag might be useful in tuning the process control algorithm. Absence of a lag between two parameters implies that the relation between the parameters depends on factors that are not correlated during the entire torrefaction run.

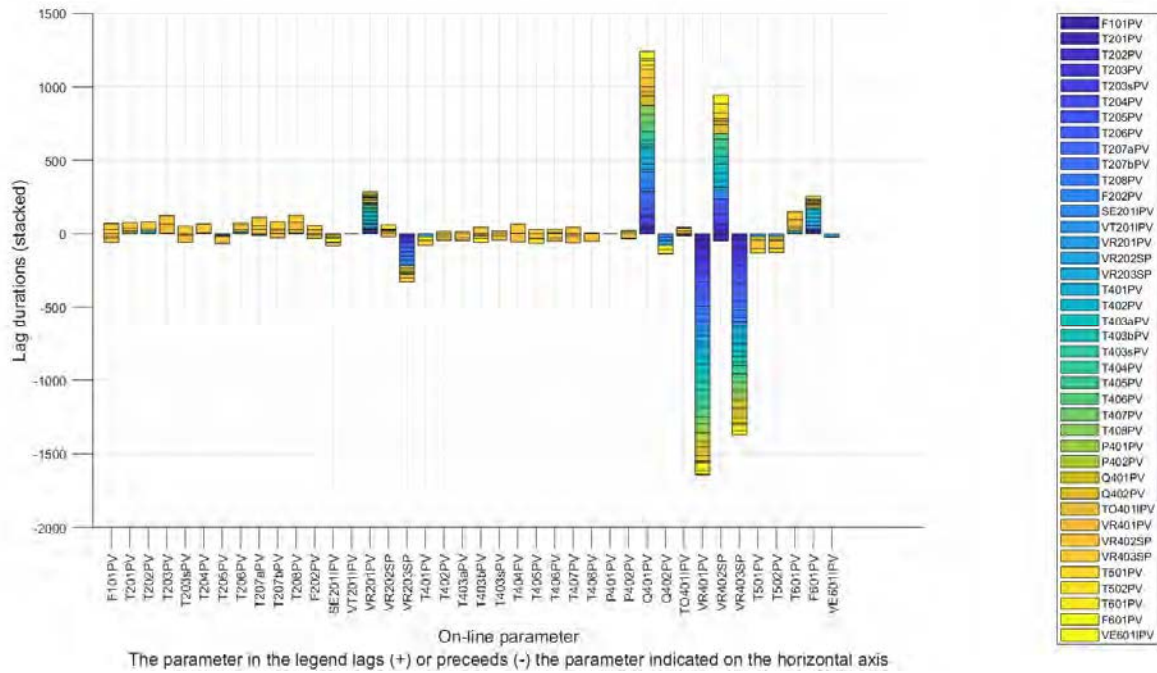


Figure 12 Lag duration

As shown in figure 12, the thermal oil control valves of both the dryer and the torrefactor are correlated with many other process parameters. The O₂ content of the torrefaction gas is also correlated with many other parameters. The average lag time between these parameters and the group of other correlated parameters is presented in table 3.

| Parameter | mean lag (minutes) |
|-----------|--------------------|
| VR201PV | -6.73 |
| VR202SP | -0.87 |
| VR203SP | 6.88 |
| VR401PV | 41.1 |
| VR402SP | -23.58 |
| VR403SP | 34.26 |
| Q401PV | -31.02 |

Table 3 Mean lag times relative to the on-line parameter

5 CONCLUSIONS

This study showed that PCA and PLS are useful methods for analyzing multivariate data obtained from a biomass torrefaction system. PCA of the entire system during stable operation showed that the temperature of the final dryer section, the temperature of the chips leaving the dryer and the temperature of the hot gas leaving the burner make the greatest contribution to the overall variance in the operational parameters. Principal component score plots obtained during periods of process instability are particularly useful in identifying the parameters that contribute the most to the total variation. The temperature of the thermal oil entering the dryer and the temperature of the gas leaving the dryer are reliable indicators of a disturbed process. PLS identified the predictor variables that make the most important contribution to the product calorific value. The temperature in the central section and the position of the valve controlling thermal oil flow to the central section of the torrefactor in particular are important parameters that can be manipulated in order to increase the product calorific value. An equation that estimates the product calorific value as a function of 40 predictor variables reduced to 9 principal components was developed and cross validated with data obtained from a separate torrefaction batch. The valves that control thermal oil flow to the different sections of the dryer and the

torrefactor were found to have lagged correlations with many other measured parameters. These lag times should be accounted for in the process control algorithm.

11 ACKNOWLEDGEMENTS

The author thanks the canton of Vaud, Switzerland for funding the torrefaction project within the framework of the program "100 million for renewable energy and energy efficiency", GRT Technologies and Engineering (GRT) SA, Orbe, Switzerland for design, construction and operation of the torrefaction pilot plant, SCCER Biosweet [5] for supporting the study of torrefaction in the context of biomass energy research and Prof. Jean-Bernard Michel for leading torrefaction activities at the HES-SO, HEIG-VD campus.

12 REFERENCES

- [1] MacGregor, J.F., Kourti, T.(1995). "Statistical Process Control of Multivariate Processes", Control Engineering Practice, Vol 3, pp. 403-414.
- [2] Eriksson, E., Johansson, E., Kattaneh-Wold, N., and S. Wold (2001), Multi-and Megavariate Data Analysis, Principles and Applications, Umetrics AB, Umeå, SE.
- [3] Wold, S., Sjöström, M., Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems vol 58. Pp 109–130.
- [4] Jackson, J.E., (1991), A User's Guide to Principal Components, John Wiley & Sons, New York, USA.
- [5] Lestander, T.A., Holmberg, C., Stenberg, L., Lehtonen, R. (2012). Towards multivariate statistical process control in the wood pellet industry. Biomass and Bioenergy, vol 45, pp 152-158.
- [6] ISO/TS 17225, Solid biofuels -- Fuel specifications and classes – Part 1: General requirements (2014) and Part 8: Graded thermally treated and densified biomass fuels (December 2016)
- [7] MATLAB R2017a version 9.2.0. Natick, Massachusetts: The MathWorks Inc., 2017.
- [8] SCCER Biosweet, Biomass for Swiss Energy Future, Swiss Competence Center for Energy Research, www.sccer-biosweet.ch



MULTIVARIATE ANALYSIS OF BIOMASS TORREFACTION PILOT PLANT OPERATION DATA TO SUPPORT THE DEVELOPMENT OF A PROCESS MODEL AND CONTROLLER

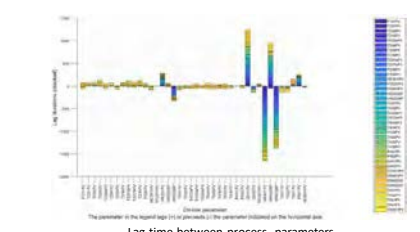
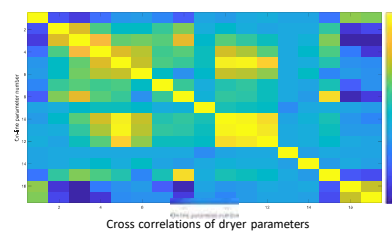
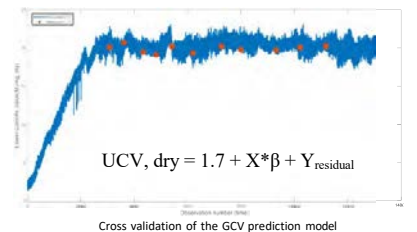
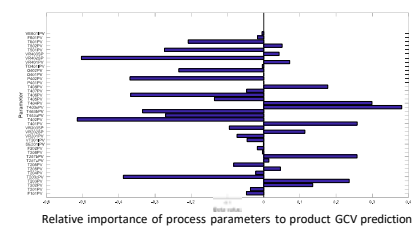
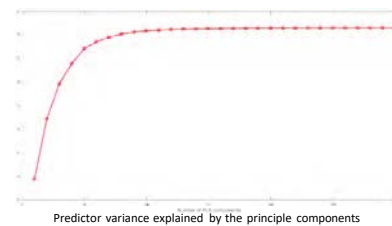
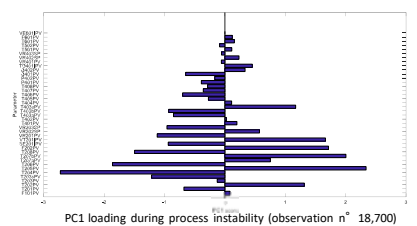
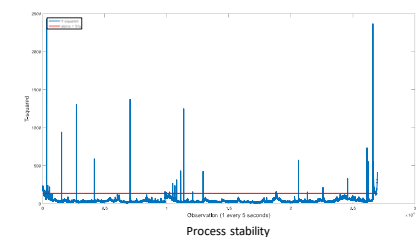
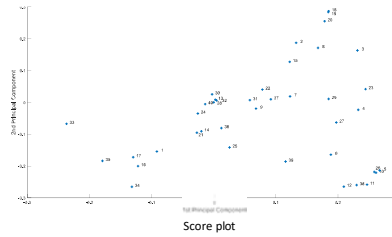
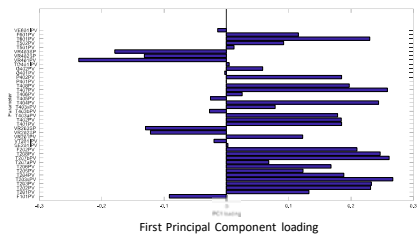
Mark McCormick¹ and Boris Correa²

1. University of Applied Sciences and Arts Western Switzerland, HEIG-VD, Thermal Engineering Institute, Industrial Bioenergy Systems unit, Centre St Roch, Avenue des sports 20, 1401 Yverdon-les-Bains, Switzerland. 2. GRT Technologies and Engineering (GRT) SA, Rue des Ducats 40 B, 1350 Orbe, Switzerland. *Corresponding author: tel: ++41 (0)24 557 61 51, E-mail: mark.mccormick@heig-vd.ch

Purpose of the study

Torrefaction of biomass is a thermal treatment process executed in an oxygen poor environment in order to improve storage and combustion properties. The purpose of the work was to gain insights into the process behavior that might be useful in the conception of a process controller. Methodical analysis of on-line and off-line data are required to identify the relevant measured and manipulated variables and to rank them in terms of their explanation of process variance and correlation thereby leading to useful insight into the process and contributing to the development of strategies to control the process. This paper reports the first Principal Components Analysis (PCA) and Partial Least Squares (PLS) regression analysis of the pilot plant operation data.

| Probe ID | Parameter | Param. N° | Probe ID | Parameter | Param. N° |
|----------|---|-----------|----------|--|-----------|
| F181PV | Biomass feed rate | 1 | T4038PV | Final torrefactor heating segment | 21 |
| T201PV | Initial dryer heating segment | 2 | T4039PV | Torrefactor final chamber | 22 |
| T202PV | Middle dryer heating segment | 3 | T4040PV | Thermal oil temperature, torrefactor | 23 |
| T203PV | Final dryer heating segment | 4 | T4050PV | Immersed in wood chips, torrefactor exit | 24 |
| T204PV | Final dryer heating segment | 5 | T4060PV | Torrefaction gas, burner entrance | 25 |
| T205PV | Dryer thermal oil | 6 | T4070PV | Exhaust gas, burner exit | 26 |
| T206PV | Post dryer exit gas | 7 | T4080PV | Exhaust gas, between the heat exchangers | 27 |
| T207PV | Heat exchanger exit gas | 8 | T4090PV | Torrefaction gas pressure | 28 |
| T208PV | Wood surface, dryer exit | 9 | T420PV | Thermal oil reservoir pressure, torrefactor | 29 |
| T209PV | Immersed in wood chips, dryer exit | 10 | Q401PV | O ₂ torrefaction gas | 30 |
| T210PV | Hot gas entering dryer | 11 | Q402PV | CO ₂ torrefaction gas | 31 |
| T211PV | Gas flow rate entering dryer | 12 | T3003PV | Current, torrefactor motor | 32 |
| S2011PV | Current, dryer motor | 13 | VR402PV | Thermal oil by-pass valve, torrefactor | 33 |
| VT001PV | Current, rotary valve between the dryer and torrefactor | 14 | VR402SP | Thermal oil by-pass valve, torrefactor central section | 34 |
| VR001PV | Thermal oil by-pass valve, dryer | 15 | VR403SP | Thermal oil by-pass valve, torrefactor final section | 35 |
| VR002SP | Thermal oil by-pass valve, dryer central section | 16 | T501PV | Interior cooling screw, midway | 36 |
| VR003SP | Thermal oil by-pass valve, dryer final section | 17 | T502PV | Interior cooling screw, exit | 37 |
| T601PV | Initial torrefactor heating segment | 18 | T602PV | Turbo exit | 38 |
| T603PV | Middle torrefactor heating segment | 19 | F901PV | Methane flow rate | 39 |
| T604PV | Final torrefactor heating segment | 20 | VE601PV | Current, dryer ventilator motor | 40 |



Conclusions

- The temperature of the thermal oil entering the dryer and the temperature of the gas leaving the dryer are reliable indicators of a disturbed process.
- The temperature in the central section and the position of the valve controlling thermal oil flow to the central section of the torrefactor in particular are important predictors that can be manipulated in order to increase and maintain the product calorific value.
- There is a lag time of 20 to 35 minutes between changes in thermal oil temperature in the torrefactor and many other process parameters.



LSTM and 1-D Convolutional Neural Networks for Predictive Monitoring of the Anaerobic Digestion Process

Mark McCormick^(✉) and Alessandro E. P. Villa

NeuroHeuristic Research Group, University of Lausanne, Quartier UNIL-Dorigny,
1015 Lausanne, Switzerland
mark.mccormick@unil.ch,
<http://www.neuroheuristic.org>

Abstract. Anaerobic digestion is a natural process that transforms organic substrates to methane and other products. Under controlled conditions the process has been widely applied to manage organic wastes. Improvements in process control are expected to lead to improvements in the technical and economic efficiency of the process. This paper presents and compares 3 different neural network model architectures for use as anaerobic digestion process predictive models. The models predict the future biogas production trend from measured physical and chemical parameters. The first model features an LSTM layer, the second model features a 1-D convolutional layer and the third model combines 2 separate inputs and parallel treatment using LSTM and 1-D convolutional layers followed by merging to produce a single prediction. The predictions can be used to adaptively adjust the substrate feeding rate in accordance with the transient state of the digestion process as defined by liquid feeding rate, the organic acid and ammonium ion concentrations and the pH of the digester liquid phase. The training and testing data were obtained during 1 year of continuous operation of a pilot-plant treating restaurant wastes. PLS regression and ICA were used to select the most relevant process parameters from the data. The 1-D Convolutional based model comprising 272 trainable parameters predicted the future biogas flow rate changes with accuracy as high as 89% and an average accuracy of 58%. The work-flow can be applied to optimize the control of the study digester and to control bioreactors in general.

Keywords: LSTM, 1-D Convolutional Neural Network · Bioprocess

1 Introduction

Restaurants, hotels, markets, fisheries and other small to medium size agro-food industries generate 88 million tonnes of organic waste per year in Europe of which 47 million tonnes per year are household food wastes and 17 million tonnes per year are food processing wastes [1]. This waste stream contains valuable components such as water (approximately 80% by mass) and valuable energy and

© The Author(s) 2019

I. V. Tetko et al. (Eds.): ICANN 2019, LNCS 11731, pp. 725–736, 2019.

https://doi.org/10.1007/978-3-030-30493-5_65

substances such as proteins and organic acids. In the context of the circular economy and current objectives to improve energy efficiency, recycle water and valuable substances, produce energy from renewable sources, reduce greenhouse gas emissions and close nutrient cycles, the waste management sector offers many opportunities to implement innovative technical solutions. Anaerobic digestion has been proposed as a technical solution to manage organic wastes. This study aims to contribute to the strategy used to control a high-performance anaerobic digester supplied by Digesto Sarl. The strength of the Digesto® digestion module is the ability to treat small to medium quantities of waste on their site of production, without any transport and in an automated mode (programmable logic controller for distance monitoring of the digester performances and for breakdown prevention). The elimination of transport costs and a high degree of automation make it the appropriate treatment solution for many end-users [2].

Currently, anaerobic digestion is successfully applied to reduce the organic matter content and to improve the dewatering properties of excess wastewater treatment sludge. Anaerobic digestion is also used to produce energy from municipal solid wastes, garden wastes and from energy crops. These processes are characterized by large bioreactors, long hydraulic retention times between 20 and 60 days and by regular feeding of well characterized and homogeneous substrates. In contrast, small-scale, on-site digestion of restaurant wastes would require an intensified bioprocess featuring a small digester, a short hydraulic retention time of less than 10 days and the capacity to tolerate wide variations in the feeding rate and the substrate composition.

Anaerobic digestion is a complex process characterized by non-linear functions and interactions between many different biochemical processes. A high-performance control strategy is thus required to intensify the anaerobic digestion process and to achieve the goals of small size, a short hydraulic retention time and long-duration autonomous operation. Artificial Neural Networks, and especially Deep Neural Networks, have demonstrated their capacity to map complex and non-linear relations between data. The purpose of this work is to construct models based on some popular Neural Network architectures and to assess their capacity to predict biogas production.

2 Materials and Methods

2.1 Machine Learning

Multivariate statistical analysis approaches have been used to detect faults and abnormal operation of anaerobic digestors treating waste activated sludge (WAS). Using Principal Component Analysis (PCA), Hotelling's T-squared and Shewhart control charts, the transitions to unstable periods were associated with accumulation of volatile fatty acids [3]. Computational self-adapting methods (Support Vector Machines, SVM) were compared with an analytical method to predict the total ammonia nitrogen (TAN) concentration in the effluent from a

two-stage anaerobic digestion (AD) process treating poultry wastes. The SVM-based model outperformed the analytical method for the TAN prediction, achieving a relative average error of 15.2% against 43% for the analytical method. Moreover, SVM showed higher prediction accuracy in comparison with Artificial Neural Networks [4]. Machine learning methods have been used to estimate the state of biogas plants, as defined by the ADM-1 model of the anaerobic digestion process, using on-line measurements of parameters such as biogas production, CH_4 and CO_2 content in the biogas, pH value and substrate feed volume [5].

In contrast to the conventional approach of using recurrent networks, especially LSTM layers, for sequence processing, convolutional architectures have demonstrated longer effective memory and have outperformed recurrent networks across a diverse range of tasks [6]. In the field of aircraft control, a “state-image” approach to capture the in-flight state variables produced a feature map that contained the values of the in-flight parameters at a given moment. The control strategy also included historical data. A Convolutional Neural Network (CNN) and a Recursive Neural Network with Long Short-Term Memory (RNN (LSTM)) layers were implemented in parallel to output vectors that were subsequently merged and processed in fully connected layers. Inputting data to different branches to allow separate extraction of the time dependent and the current parameter values was expected to lead to improved accuracy when compared to sequential treatment by CNN and LSTM layers [7].

2.2 Experimental Set-Up

The study digester was a machine comprising tanks, pumps, heating elements, sensors and command and data acquisition capacities. The total liquid working volume was 630 liters. Shredded restaurant wastes were fed to the digester at regular intervals. The study data was acquired from on-line measurements and by analysis of samples collected at intervals between 1 and 7 days during 366 days of continuous operation. The study data included 22 parameters that characterize the input waste stream. Characterization of the digestion process included manual measurement of 44 parameters that describe the liquid and gas phases. The on-line measurements included redundant measurements of liquid and gas phase physical-chemical properties and the measurement of 7 mechanical parameters such as internal pumping and agitation. A total of 83 parameters were considered for the study. The digestion process was controlled manually using only the expert knowledge of the operators. The operating conditions were adjusted *ad hoc* during the study to optimize the rates of biogas production and the removal of the organic fraction of the waste. In this study, fixed duration feeding intervals were defined. At the start of each time interval a decision was made to feed or to not feed the substrate to the digestion machine.

2.3 Problem Solving Approach

Autonomous operation of the digester would require automated transfer of raw waste from a buffer tank to the digestion machine. The problem is to decide when

the raw waste should be transferred (fed) to the digestion machine. Overfeeding would upset the bioprocess and can lead to ceased waste degradation and biogas production. Underfeeding would lead to inefficient use of the digester. A useful model should make predictions that agree with the actual values that were measured after the prediction. The digester should be fed only when the biological state is conducive to waste degradation and biogas production. Therefore, the purpose of the neural network model is to automate decision making. To solve the problem, a supervised approach based on mapping measured process predictor variables to a single measured response variable was chosen. The models were trained and tested using simultaneously measured prediction and response values. The models were tested by comparing predicted to measured values of the response variable.

The proposed anaerobic digester control strategy is to compare the predicted future biogas flow rate trend to the actual biogas flow rate TREND. Increased biogas production rate is associated with improved bioprocess quality. Decreased or stable biogas production rate is associated with deteriorating bioprocess quality. If the predicted future biogas flow rate is higher than the actual measured biogas flow rate, then the decision is to feed the digestion machine. If the predicted biogas flow rate was the same as or lower than the actual measured biogas flow rate, then the decision was to not feed the digestion machine.

2.4 Data Analysis and Preprocessing

Analysis and data augmentation techniques were used to prepare a single dataset for use in developing all the Neural Network models. Raw data analysis aimed to identify the parameters to use to train and test the Neural Network model. Principal Component Analysis (PCA), Independent Component Analysis (FastICA) and Partial Least Squares (PLS) regression algorithms from Scikit-learn [8] were used for dimensionality reduction and to identify the parameters that make the greatest contribution to process stability. A plot of the weights of the first principal components shows that the waste loading rate, total solids, carbon and nitrogen content of the feed, and the total solids concentration in the of the digester make the greatest contributions to overall variance of the system. However, these parameters can be measured only manually and manipulation by the operators is difficult or impossible. Consequently, they are not practically useful in a neural network model for process control.

ICA is recommended when the data has a high degree of kurtosis. Since sharp peaks were observed for biogas H_2 content, and liquid phase ammonium and volatile fatty acid content, the raw data were analyzed using ICA. ICA analysis showed that the H_2 content of the biogas and the volatile fatty acid content of the digester were the most important independent components of the data. The raw data was used in the DNN models without pre-treatment to remove kurtosis because spiking of both H_2 and volatile fatty acids is not unusual in the anaerobic digestion process. The H_2 concentration in the biogas can be measured on-line. However, no sensors for volatile fatty acids are currently available at an affordable price. The method of Partial Least Squares Regression was used to

map the matrix of predictor variables to the biogas production rate. The results of PLS analysis showed that regression equation coefficients for the nitrogen content of the feed and the digester loading rate had the greatest magnitudes.

Considering the results of PCA, ICA and PLS regression analysis, knowledge of the anaerobic digestion process and the practical aspects of data acquisition, the following parameters were selected for inclusion in the dataset used to build the Neural Network models with 4 input features and 1 target variable.

- *Predictor variables (input features)*
 - Total mass of feed (water + waste + co-substrate) [kg/day]
 - pH
 - Ammonium concentration [mg N/liter]
 - Volatile fatty acid concentration [mg Acetic Acid/liter]
- *Response variable (target)*
 - Biogas flow rate [liter/hour]

Pre-processing included augmentation of the number of data points. In particular, the frequency of values obtained from off-line measurements of physical-chemical parameters was increased from approximately 1 per week to 1 per hour. The number of data points was increased by resampling to create one-hour time intervals followed by linear interpolation of the off-line data to fill the new sampling times. Interpolation was assumed to be valid because the changes in biological systems occur slowly over several days. The augmented dataset included 8733 rows of data representing the selected input features and the target. Using the Scikit-learn MinMaxScaler function, the previously standardized data were then normalized to values between 0 and 1. Figure 1 shows the pre-processed dataset that was used to build the Neural Network models. Five series of training and test data were obtained using the Scikit-learn TimeSeriesSplit function which respects the sequential order of the original data.

2.5 Neural Network Architectures

The study aimed to compare the recurrent, convolutional and combined neural network modeling approaches to solving a time series analysis problem having at 4 features. The models were built using Python 3.6.8 and the Keras 2.2.4 API running on top of TensorFlow 1.3.0 library for machine learning. The 3 architectures selected for development and the rationale for selection are summarized in Table 1 and the models are summarized in Fig. 2.

Long Short-Term Memory (LSTM). The sequential model was constructed using a single LSTM layer followed by a Dense layer. The input was a 3D tensor with shape (`batch_size`, `timesteps`, `input_dim`) where `batch_size` is the number of sample batches, `timestep` was set to four hours and `input_dim` was equal to the number of features. The LSTM layer had 160 units, used the `tanh` activation and `hard_sigmoid` recurrent activation functions with zero dropout. `Unit_forget_bias` was set to false since this setting was found to give a more

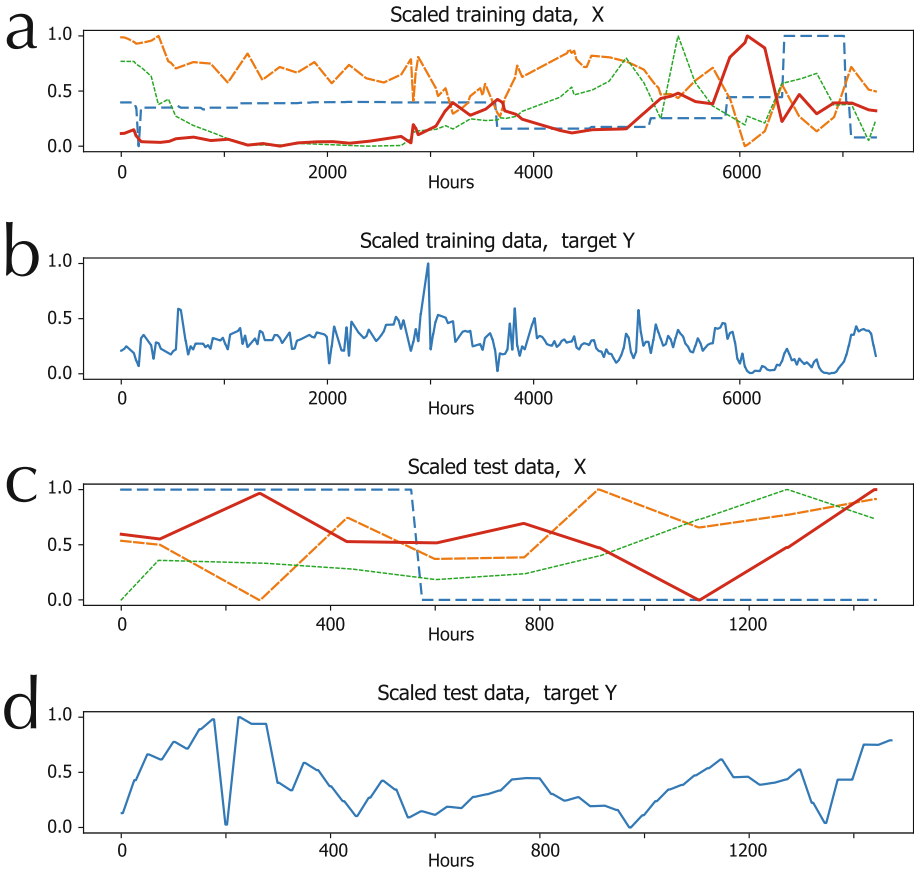


Fig. 1. Training and testing data after preprocessing. **(a)** Normalized training data of predictor variables. blue line: total volume of feed; red line: pH; green curve: ammonium concentration; orange line: volatile fatty acid concentration. **(b)** Normalized training data of the response variable, i.e. biogas flow rate. **(c)** Normalized test data of predictor variables. Same color code of panel (a). **(d)** Normalized test data of the response variable. (Color figure online)

accurate prediction. The model output was from a single dense layer having 1 unit, `relu` activation and `use_bias` set to `False`. The model had 105,760 trainable parameters.

1-Dimensional Convolutional Layer (Conv1-D). The sequential model was constructed using a single `Conv1D` layer followed by a `MaxPooling1D` layer followed by a `Dense` layer. The input was a 3D tensor with shape `(batch_size, timesteps, input_dim)`, where `batch_size` is the number of sample batches, `timesteps` was set to four hours and `input_dim` was equal to the number of features. The `Conv1D` layer had 16 filters, a kernel size of 4, stride of 1, padding

Table 1. Evaluated Neural Network architectures

| Neural network architecture | Rationale for selection |
|---|---|
| LSTM, biogas flow rate predictions | Expert knowledge that bioprocesses have long and variable lag-times between measured predictor values and responses. Historical information |
| 1-D CNN, biogas flow rate predictions | Potential to extract useful information from detailed analysis of features at given moments. State image concept |
| LSTM/1-D CNN hybrid, biogas flow rate predictions | Potential to improve accuracy by separate processing of historical and state image information followed by merging the information |

set to same, and used the `relu` activation function. The pool size of the Max-Pooling layer was set to 4. The model output was from a single dense layer having 1 unit, `relu` activation and `use_bias` set to False. The model had 272 trainable parameters.

LSTM/Conv1-D Hybrid. The Keras Model Class API was used to construct a model having 2 separate input and processing branches and a single output. The first branch included an LSTM layer followed 2 dense layers. The LSTM layer had the same configuration as the layer described above. The first dense layer had 48 nodes and used the `relu` activation function. The final dense layer had 1 node and used the `relu` activation function. The second branch included a Conv1D layer followed by MaxPooling1D Dense and Flatten layers. The Conv1D layer had the same configuration as the layer described above. The MaxPooling1D layer strides were set to 1 and padding was set to same. The two branches were merged using the Keras concatenate layer. The merged output was further passed to a dense layer with 8 hidden nodes. Model output was from a dense layer having 1 node and using the `relu` activation. The model had 106,158 trainable parameters.

3 Results

The augmented dataset was split into 5 different pairs of training and test sets using the `TimeSeriesSplit` function. Each pair comprised separate sets of consecutive 1-h intervals and a total of between 5000 and 8500 timestamps. The LSTM, Conv1D and hybrid LSTM/Conv1D models were trained for 1000 epochs with a batch size of 4 h. Loss during training was assessed using mean square error. The loss during training of the 3 models is shown in Fig. 3.

The models were tested on data obtained previously using the `TimeSeriesSplit` function. The prediction accuracy was assessed visually by plotting the predicted values and the measured values at 4-h intervals corresponding to the

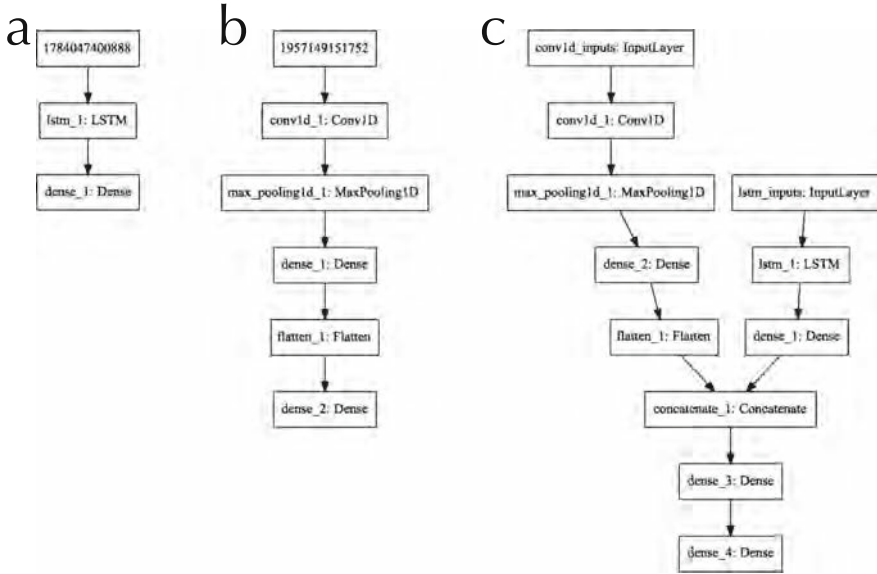


Fig. 2. Artificial Neural Network architectures: (a) LSTM, (b) Conv1D and (c) Hybrid LSTM/Conv1D architectures

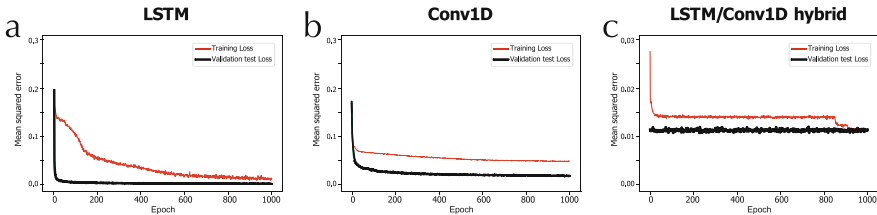


Fig. 3. Comparison of training loss for the 3 ANN architectures.

batch-size of 4 times 1-h timestamps. The RMSE for the individual sample pairs was calculated at the same 4-h interval (batch).

The predicted and the observed biogas flow rates for the 3 different regression models are shown in Fig. 4abc. The RMSE of each batch is shown in Fig. 4def. To assess the accuracy of the 3 models over a long duration, the overall RMSE was also calculated for the entire population of batches.

Table 2 shows the overall RMSE of the 3 models.

Since the purpose of the model is to decide if the digester should be fed at the beginning of a time interval, a special evaluation procedure was developed to rate the models in terms of their capacity to make the correct decision regarding digester feeding. To rate the models, the consequences of the model’s decision were compared to the actual change in measured biogas flow rate 4 h after the decision. Any decision to feed the digester, followed by an observed

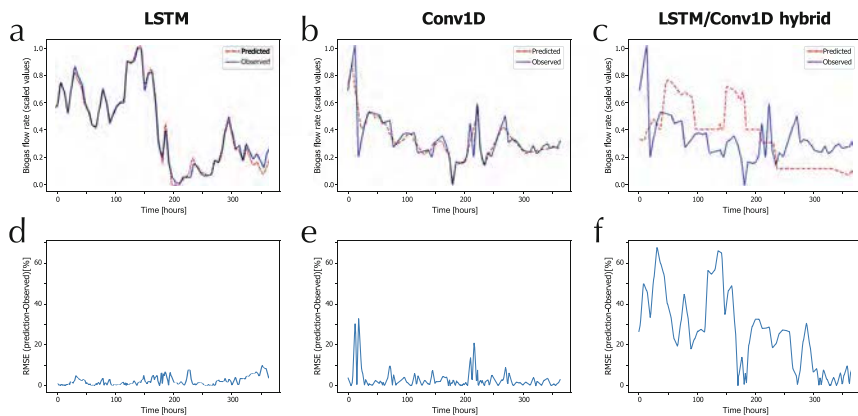


Fig. 4. Predicted (dashed red curve) and observed (solid blue curve) biogas production for LSTM (panel a), 1-D CNN (panel b), and LSTM/1-D CNN hybrid (panel c). The corresponding errors (RMSE) are presented in panels d,e,f for LSTM, 1-D CNN and LSTM/1-D CNN hybrid ANN architectures, respectively. (Color figure online)

Table 2. RMSE of the entire population

| Model (ANN architecture) | RMSE (%) |
|--------------------------|----------|
| LSTM | 1.7 |
| 1-D CNN | 2.6 |
| LSTM/1-D CNN hybrid | 19.0 |

decrease in biogas production, was an incorrect decision. Any decision to feed the digester, followed by an observed stable or increased biogas production, was a correct decision. The model accuracy was simply the ratio of correct decisions to requested decisions expressed as a percentage. To validate the 1-D convolutional DNN model, 12 new data sets generated by splitting the original data were used to predict the biogas production trend 4-h in the future. The average accuracy of the prediction was 58% with a standard deviation of 24% and a range between 28% and 100%.

To investigate the response time between changes in the measured predictor and response variables, the error of the LSTM model predictions was evaluated at different lag-times after the feeding decision. In this case, error was defined as the RMSE of the measurements and predictions made at 1-h intervals during the time sequence. Population RMSE was calculated for lag-times from 1 to 100 h after a feeding decision. The results show that the model is most accurate when the lag-time between the feeding decision and the time off comparison of the predicted to the observed values is much greater than 4 h. This result suggests that the biological response time of biogas production to changes in the 4 measured process parameters used this study is longer than 4 h (Fig. 5) .

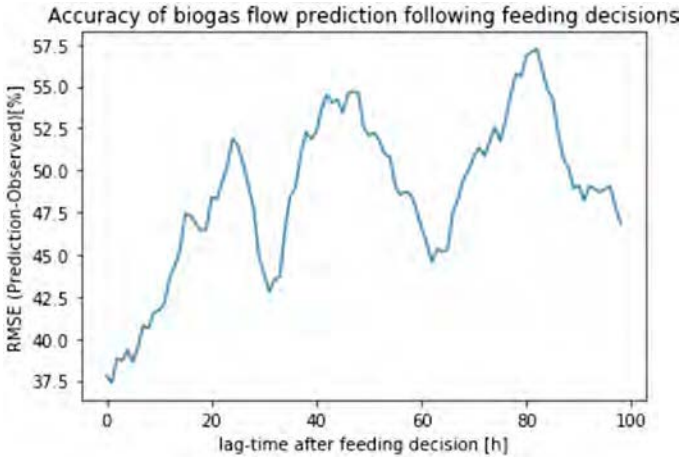


Fig. 5. RMSE of the predicted and observed biogas production at different times after measurement of predictor variables (LSTM model).

4 Discussion

The treatment of restaurant wastes is a particularly challenging application of anaerobic digestion because the loading rate and composition of the wastes vary over a relatively large range when compared to the treatment conditions for anaerobic digestion of waste activated sludge, manure and energy crops. Consequently, the feed controller must have the capacity to map a wide range of values of the relevant process parameters and account for long-duration trends. The results of this work show that a one-dimensional convolutional Deep Neural Network model trained using state image data obtained by measuring 4 parameters could be implemented as a control model to regulate substrate feeding to an anaerobic digester treating restaurant wastes. This results suggests that the convolutional network and state image approach leads to a more effective model because this approach is less influenced by the long and variable lag-times observed in a real bioprocess. In contrast, the LSTM model takes long range trends into account. Both the 1-D convolutional model and the LSTM models achieved low RMSEs between predicted and observed values of the first data set. However, model testing with new data showed that the LSTM did not have the ability to generalise. The hybrid LSTM/1-D convolutional network had the highest RMSE and the least ability to generalise.

The study demonstrated that the prediction accuracy of the neural network model is similar to the prediction accuracy of simply adjusting the feeding rate according to the observed trend in biogas production.

The average prediction accuracy of 58% should be compared to the accuracy of 63% obtained by simply following the trend in biogas production. The trend in biogas production was the result of *ad hoc* control of the digester by skilled operators working in a laboratory. It seems unlikely that a prediction accuracy

as high as 63% could have been achieved in an industrial setting without control by skilled operators. Importantly, the measured parameters of input liquid flow, output biogas flow, digester pH and digester ammonium content can be measured on-line using commercially available sensors. However, no sensor to measure volatile fatty acids is available commercially.

Considering that approximately 80% of the data points for pH, ammonium and volatile fatty acids concentrations were obtained by interpolation between values measured only 1 or 2 times per week, the accuracy of the model could be improved by increasing the sampling frequency thereby obtaining real data. Higher data resolution and consequently higher model accuracy could be achieved using on-line sensors.

5 Conclusion

This study evaluated 3 Neural Network architectures for use as supervised learning models to predict biogas production in an anaerobic digester. The models were constructed, trained and tested on the same time series dataset obtained from long-duration operation of an anaerobic digester treating restaurant wastes. Of the 3 models evaluated, the 1-D convolutional model was best able to accurately predict biogas production trends on new data as evaluated in terms of the ratio of correct predictions to the total number of requested predictions. The results suggest that the feeding rate of the study digester can be controlled using an 1-D Convolutional based DNN controller. Future work should aim to improve the resolution of the training data set by implementing on-line sensing of the relevant process parameters. During this study, the digester feeding rate was the only manipulated parameter. Future work should also aim to control additional manipulatable parameters and the mechanical parameters of the digestion process. Considering the wide range of variation observed in the input and target variables, the collection of time series data over a very long duration is required to make an accurate model.

Acknowledgement. This work was supported by Swiss Innovation cheque #32609 and by F.R. Mahrer of Digesto Sarl, Puplinge(GE), Switzerland.

References

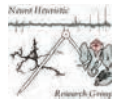
1. Stenmarck, Å., et al.: FUSIONS: estimates of European food waste levels, Stockholm (2016). ISBN 978-91-88319-01-2
2. Capaccioli, S., et al.: Socio-economic benefits of small anaerobic digestion systems. In: Proceeding of the 24th European Biomass Conference and Exhibition, pp. 1530–1534, Amsterdam (2016)
3. Leite, W.R.M., Belli Filho, P., Gottardo, M., Pavan, P., Bolzonella, D.: Monitoring and control improvement of single and two stage thermophilic sludge digestion through multivariate analysis. *Waste Biomass Valorization* **9**(6), 985–994 (2016)

4. Alejo, L., Atkinson, J., Guzmán-Fierro, V., Roeckel, M.: Effluent composition prediction of a two-stage anaerobic digestion process: machine learning and stoichiometry techniques. *Environ. Sci. Pollut. Res.* **25**(21), 21149–21163 (2018)
5. Gaida, D., et al.: State estimation for anaerobic digesters using the ADM1. *Water Sci. Technol.* **66**(5), 1088–1095 (2012)
6. Bai, S., Kolter, J.Z., Koltun, V.: An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. [arXiv: e1803.01271](https://arxiv.org/abs/1803.01271) (2018)
7. Dong, Y.: An application of deep neural networks to the in-flight parameter identification for detection and characterization of aircraft icing. *Aerosp. Sci. Technol.* **77**, 34–49 (2018)
8. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**(Oct), 2825–2830 (2011)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





The Problem

Restaurants, hotels, markets, fisheries and other small to medium size agro-food industries generate 88 million tonnes of organic waste per year in Europe of which 47 million tonnes per year are household food wastes and 17 million tonnes per year are food processing wastes. Anaerobic digestion has been proposed as a technical solution to manage organic wastes. Currently, anaerobic digestion is successfully applied to reduce the organic matter content and to improve the dewatering properties of excess wastewater treatment sludge. Anaerobic digestion is also used to produce energy from municipal solid wastes, garden wastes and from energy crops. These processes are characterized by large bioreactors, long hydraulic retention times between 20 and 60 days and by regular feeding of well characterized and homogeneous substrates. In contrast, small-scale, on-site digestion of restaurant wastes would require an intensified bioprocess featuring a small digester, a short hydraulic retention time of less than 10 days and the capacity to tolerate wide variations in the feeding rate and the substrate composition.

The problem is to decide when the raw waste should be transferred (fed) to the digestion machine. A useful model should make predictions that agree with the actual values that were measured after the prediction. The purpose of this work is to construct models based on some popular Artificial Neural Network (ANN) architectures and to assess their capacity to predict biogas production with the aim of using the models in anaerobic digester process control.

Methods

The study digester was a machine comprising tanks, pumps, heating elements, sensors and command and data acquisition capacities. The study data was acquired from on-line measurements and by analysis of samples collected at intervals between 1 and 7 days during 366 days of continuous operation. A total of 83 parameters were considered for the study. The digestion process was controlled manually using only the expert knowledge of the operators. The operating conditions were adjusted *ad hoc* during the study to optimize the rates of biogas production and the removal of the organic fraction of the waste.

Considering the results of PCA, ICA and PLS regression analysis, knowledge of the anaerobic digestion process and the practical aspects of data acquisition, the following parameters were selected for inclusion in the dataset used to build the Neural Network models with 4 input features and 1 target variable (see Table 1).

Table 1. Parameters of the model

| Predictor variables (input features) | Response variable (target) |
|--|--------------------------------|
| Total mass of feed (water + waste + co-substrate) [kg/day] | Biogas flow rate [liters/hour] |
| pH | |
| Ammonium concentration [mg N/liter] | |
| Volatile fatty acid concentration [mg Acetic Acid/liter] | |

The number of data points was increased by resampling to create one-hour time intervals followed by linear interpolation of the on-line data to fill the new sampling times. Interpolation was assumed to be valid because the changes in biological systems occur slowly over several days. The augmented dataset included 8733 rows of data representing the selected input features and the target. Using the Scikit-learn MinMaxScaler function, the previously standardized data were then normalized to values between 0 and 1. Five series of training and test data were obtained using the Scikit-learn TimeSeriesSplit function which respects the sequential order of the original data.

The study aimed to compare the recurrent, convolutional and combined neural network modeling approaches to solving a time series analysis problem having 4 features. The 3 architectures selected for development and the rationale for selection are summarized in Table 2 and the models are summarized in Figure 1.

Table 2. Evaluated Neural Network architectures

| Neural Network architecture | Rationale for selection |
|-----------------------------|--|
| LSTM | Expert knowledge that bioprocesses have long and variable lag-times between measured predictor values and responses. Capacity to extract historical information. |
| 1-D CNN, | Potential to extract useful information from detailed analysis of features at given moments. State image concept. |
| LSTM/1-D CNN hybrid, | Potential to improve accuracy by separate processing of historical and state image information followed by merging the information. |

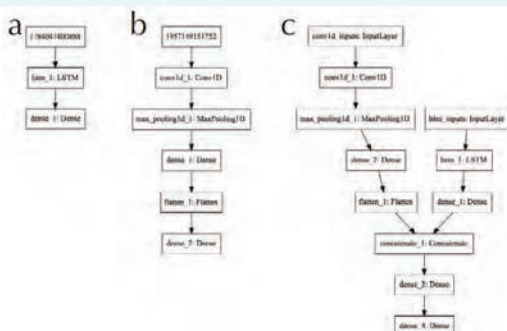


Figure 1. Artificial Neural Network architectures: (a) LSTM, (b) Conv1D and (c) Hybrid LSTM/Conv1D

Results

The LSTM, Conv1D and hybrid LSTM/Conv1D models were trained for 1000 epochs with a batch size of 4 hours. Loss during training was evaluated using mean square error. The loss during training of the 3 models is shown in Figure 2. No significant improvement in the training loss was expected after 1000 epochs.

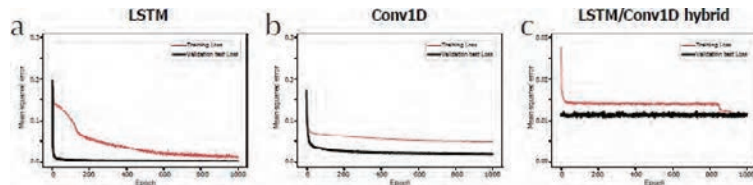


Figure 2. Comparison of training loss for the 3 ANN architectures.

The predicted and the observed biogas flow rates for the 3 different regression models are shown in Figure 3abc. The RMSE of each batch is shown in Figure 3def. To assess the accuracy of the 3 models over a long duration, the overall RMSE was also calculated for the entire population of batches. The resulting RMSEs of the LSTM, 1-D CNN and the LSTM/1-D CNN hybrid were respectively 1.7, 2.6 and 19.0%.

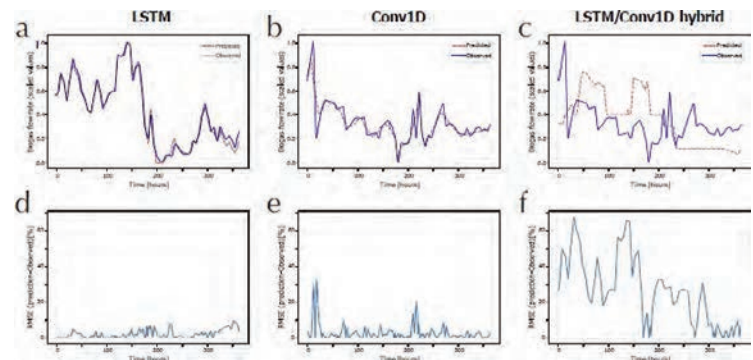


Figure 3. Predicted (dashed red curve) and observed (solid blue curve) biogas production for LSTM (panel a), 1-D CNN (panel b), and LSTM/1-D CNN hybrid (panel c). The corresponding errors (RMSE) are presented in panels d,e,f for LSTM, 1-D CNN and LSTM/1-D CNN hybrid ANN architectures, respectively.

Since the purpose of the model is to decide if the digester should be fed at the beginning of a time interval, a special evaluation procedure was developed to rate the models in terms of their capacity to make the correct decision regarding digester feeding. To rate the models, the consequences of the model's decision were compared to the actual change in measured biogas flow rate 4 hours after the decision. Any decision to feed the digester, followed by an observed decrease in biogas production, was an incorrect decision. Any decision to feed the digester, followed by an observed stable or increased biogas production, was a correct decision. The model accuracy was simply the ratio of correct decisions to requested decisions expressed as a percentage. Only the 1-D convolutional ANN demonstrated the capability to generalize. To further validate the 1-D convolutional ANN model, 12 new data sets generated by splitting the original data were used to predict the biogas production trend 4-hours in the future. The average accuracy of the prediction was 58 % with a standard deviation of 24 % and a range between 28 % and 100 %.

Discussion

The results of this work show that a 1-dimensional convolutional Deep Neural Network model trained using state image data obtained by measuring 4 parameters could be implemented as a control model to regulate substrate feeding to an anaerobic digester treating restaurant wastes. This results suggests that the convolutional network and state image approach leads to a more effective model because this approach is less influenced by the long and variable lag-times observed in a real bioprocess. In contrast, the LSTM model takes long range trends into account. Both the 1-D convolutional model and the LSTM models achieved low RMSEs between predicted and observed values of the test data set. However, model testing with new data showed that the LSTM did not have the capability to generalize. The hybrid LSTM/1-D convolutional network had the highest RMSE and the least capability to generalize.

The study demonstrated that the prediction accuracy of the 1-D convolutional neural network model is similar to the prediction accuracy achieved by simply adjusting the feeding rate according to the observed trend in biogas production. The average prediction accuracy of 58% should be compared to the accuracy of 63% obtained by simply following the trend in biogas production. The trend in biogas production was the result of *ad hoc* control of the digester by skilled operators working in a laboratory. It seems unlikely that a prediction accuracy as high as 63% could have been achieved in an industrial setting without control by skilled operators.

Article

An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process

Mark McCormick

Neuroheuristic Research Group, Department of Information Systems, Faculty of Business and Economics, Quartier UNIL-Chamberonne, University of Lausanne, 1015 Lausanne, Switzerland; mark.mccormick@unil.ch

Abstract: The purpose of this work was to develop a problem-solving approach and a simulation tool that is useful for the specification of wastewater treatment process equipment design parameters. The proposition of using an artificial neural network (ANN) numerical model for supervised learning of a dataset and then for process simulation on a new dataset was investigated. The effectiveness of the approach was assessed by evaluating the capacity of the model to distinguish differences in the equipment design parameters. To demonstrate the approach, a mock dataset was derived from experimentally acquired data and physical effects reported in the literature. The mock dataset comprised the influent flow rate, the bed packing material dimension, the type of packing material and the packed bed height-to-diameter ratio as predictors of the calorific value reduction. The multilayer perceptron (MLP) ANN was compared to a polynomial model. The validation test results show that the MLP model has four hidden layers, each having 256 units (nodes), accurately predicts calorific value reduction. When the model was fed previously unseen test data, the root-mean-square error (RMSE) of the predicted responses was 0.101 and the coefficient of determination (R^2) was 0.66. The results of simulation of all 125 possible combinations of the 3 mechanical parameters and identical influent wastewater flow profiles were ranked according to total calorific value reduction. A *t*-test of the difference between the mean calorific value reduction of the two highest ranked experiments showed that the means are significantly different (p -value = 0.011). Thus, the model has the capacity to distinguish differences in the equipment design parameters. Consequently, the values of the three mechanical feature parameters from the highest ranked simulated experiment are recommended for use in the design of the industrial scale upflow anaerobic filter (UAF) for wastewater treatment.

Keywords: artificial neural network; environmental engineering; wastewater treatment; upflow anaerobic filter; bioprocess simulation; anaerobic; biogas



Citation: McCormick, M. An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process. *Sustainability* **2022**, *14*, 7959. <https://doi.org/10.3390/su14137959>

Academic Editors: Xiaowei Li and José Alberto Herrera-Melián

Received: 25 May 2022

Accepted: 23 June 2022

Published: 29 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Bioprocesses are defined as controlled conditions that result in the biological agent-mediated transformation of a substrate chemical into a specific product chemical. Wastewater treatment processes are environmental bioprocesses designed to improve environmental quality. They are often designed, simulated, and optimized using the conventional International Water Association (IWA) approach to modeling the wastewater treatment process implemented as differential and algebraic equations [1]. Process specification includes the evaluation of the physical and chemical parameters of the feed and product streams and of the internal biochemical and physical-chemical process steps. Initial bioprocess development is typically conducted in small laboratory bioreactors (between 1 and 10 L working volume) of standard mechanical configurations using defined synthetic substrates under highly controlled conditions. Factorial designed experiments with polynomial regression analysis and hybrid (discrete/continuous) systems approaches are used to determine the significant parameters and the operating set points [2–4].

This conventional approach might not adequately account for the variation encountered under real industrial, and especially under real environmental conditions where it is not possible to continuously measure and regulate the substrate composition and mass flow [5,6]. In a review of the hydrodynamic models of anaerobic bioreactors, Liotta et al. [7], presented and compared the conventional numeric mathematical approaches including plug flow, tank in series, dispersion and computational fluid dynamic models. This review noted that model calibration often aims to assess key hydrodynamic parameters using tracer tests and that the kinetic constants are assessed if the biochemical modules are implemented. It was also noted that there are few attempts to apply the models for the optimum design and scale-up of bioreactors. Thus, a more comprehensive and integrated approach is needed.

Bioprocesses involve complex interactions between the temperature, liquid flow parameters, substrates and a dynamic community of microbes that are difficult to specify, model and control [8,9]. Additionally, the mechanical characteristics and constructive features of the full-scale bioreactor can be very different from those at the laboratory scale. To address these issues, many researchers have recognized the usefulness to shift towards a “black-box” approach for modeling environmental bioprocesses and engineered systems. Instead of attempting to characterize biochemical reaction mechanisms, a “black-box” approach aims to build a robust model of the process based on easily measurable system parameters such as influent flow rate, product flow, vessel dimensions and constructive features, and operating set-points [10,11]. This approach does not require the determination of bioprocess parameters such as feed and product stream composition, fluid flow parameters, kinetic coefficients, inhibition factors and the identity of the active microbes, but requires a large dataset that covers a wide range of operating conditions.

Artificial neural network (ANN) models are “black-box” models. They are calculation tools, rather than precise descriptions of chemical reaction mechanisms and physical reactors. ANNs have demonstrated their capacity to map complex and non-linear relations between data [12–14] with applications to environmental bioprocesses [15–18]. For example, an ANN model of empirical data has been developed to investigate the optimum conditions for formaldehyde removal from air by a bio trickling filter [19]. The authors used the model to assess removal efficiency as a function of operating conditions and sensitivity analysis was used to determine the relative contribution of each input parameter to the outputs. In another study, a multilayer perceptron (MLP) model was developed to simulate and predict sprinkler precipitation from the input parameters of operating pressure, wind speed, wind direction and sprinkler nozzle diameter with the goal of reducing the time and cost of determining sprinkler irrigation uniformity coefficients [20]. To optimize the hydrodynamic parameters on inverse fluidized bed reactors, an MLP model was also trained using the same data organized in the same Box–Behnken experimental plan as was used to make a model based on the response surface methodology (RSM) [21]. This study found that the ANN model more accurately predicted dependent variables whereas both the RSM and the ANN models were helpful in optimization. Moving in the same direction, another study used a macroscopic, black-box approach to overcome the difficulty in modeling a hybrid and multivariate filtration process [22]. The authors successfully approximated the breakthrough curves using an ANN model using operational parameters (i.e., feed flow rate, initial concentration, adsorption bed length and filter sheet type) as predictors, and easily measurable objective variables to evaluate the process performance. In another study, correlation analysis and decomposition of the time series were used to identify the most important process parameters [23]. The authors then built an ANN-based intelligent system to predict effluent COD from the important process parameters. The model achieved a prediction accuracy (MAPE) of 10.8%.

Anaerobic contact processes for domestic sewage treatment were first described in the mid-20th century [24–26]. In this framework, an upflow anaerobic filter (UAF) bioreactor was proposed [27–30]. The UAF is type of fixed bed bioreactor comprised of a vessel that contains a packed bed of solid supports (packing) on which a strongly attached biofilm

develops [31]. Influent wastewater enters below the bed, flows upwards, comes into contact with the biofilm and then exits the vessel from an outlet located above the bed. The biofilm is composed of a community of microorganisms which are the functional agents of the biotransformation of substrate to product. By a process known as anaerobic digestion, some components of the wastewater are transformed into biogas that exits the system, thereby reducing the water pollution. The advantage of the design is the dense, fixed film community of bacteria that tolerates variations in influent stream flow rate and quality, thus making it possible to reduce the vessel volume and to treat influent wastewater, whose characteristics such as chemical composition, temperature and flow rate vary over a wide range.

Evidently, the use of machine learning techniques is common in the field of wastewater treatment process modeling. Among machine learning techniques, it was recently reported that ANNs comprised 65% of the models used in WWTP process forecasting [23]. The majority of studies focused on wastewater treatment plant operational parameters that predict effluent quality. The models were mainly used for fault detection and operations forecasting. Very few studies have focused on mechanical design parameters as predictors of effluent quality. In this context, the present study is different because the model predictors are the mechanical features of the process equipment, and additionally, the model can be used to simulate different mechanical feature values during the equipment design process. To develop high-performance environmental bioprocesses, what is needed is a simulation tool that is capable of accounting for many complex and varying hydraulic and biological system parameters, variations over daily or even seasonal time periods, variable process control parameters, fixed mechanical constructive parameters and the non-linear and temporal relations between them [5,11,32]. Hence, a method for non-linear system identification and for fixed mechanical parameter estimation is required. This study aimed to develop a method to meet these requirements in the context of designing a fixed bed bioreactor.

The present work is a proof of concept study. The method and original workflow presented herein are intended to improve the entire environmental bioprocess development workflow from raw data collection to design specification. This study tests the hypothesis that an artificial neural network model can be used to distinguish different sets of upflow anaerobic filter mechanical set-point values by simulation. The experimental plan and the method to build and use the artificial neural network model are presented and discussed. Thus, the specific objectives of this study are to demonstrate and evaluate:

- An efficient experimental plan that can be implemented in field conditions;
- The accuracy of the artificial neural network predictive model;
- The selection of mechanical constructive parameters based on significant differences in the performance results obtained by simulation.

2. Materials and Methods

2.1. Problem Solving Approach

The approach to solving the problem consists of the following steps:

- Installation of a bioreactor testing and data acquisition system at an industrial site;
- Preparation of an experimental plan to vary the operational parameters such that the test system is exposed to the full range of possible industrial conditions;
- Operation of the test system according to the experimental plan to acquire raw data;
- Restructuring and pre-treatment of the raw dataset according to the requirements of ANN models and to describe individual experiments according to the experimental plan;
- Construction of an artificial neural network model;
- Training and validation of the ANN model using supervised learning;
- Use of the ANN model to simulate test cases where mechanical parameters are varied within the range of tested values;
- Ranking of the simulation results in terms of the predicted performance of the upflow anaerobic filter bioreactor;

- Selection of the mechanical parameters according to the ranking of the simulation results.

2.2. Direct Experimental Data

The test unit used in this study was a purpose-built cylindrical, unheated, fixed bed bioreactor having a vessel liquid volume of 37 L ($v/v = 37$ L). The test unit was installed at a municipal wastewater treatment plant located in Yverdon-les-Bains, Switzerland, and was continuously fed raw primary sedimentation basin overflow using a peristaltic pump set to deliver wastewater at a specified flow rate between 127 and 516 L·day⁻¹. Three major mechanical factors were considered to affect the bioprocess: the packing equivalent spherical diameter (ESD), the packing material type (MAT), and the fixed bed height/diameter ratio (HDR). More than 95% of the packing material used was torrefied wood chips (TWD) with low sphericity ($ESD = 12$ mm). In order to evaluate other packing materials, the remaining 5% was composed of a mixture of polyvinyl chloride (PVC) chips, polyethylene tubes, polyurethane foam (PUF) cubes, ceramic and glass chips placed throughout the fixed bed. The relative biofilm activity of each type of material was determined by the dehydrogenase enzyme activity tests [33] of packing material samples removed from the bioreactor during the experiments. The fixed bed height/diameter ratio of our experimental bioreactor was 1.8 ($HDR = 1.8$).

Data were collected during continuous operation under industrial conditions from 4 May until 12 December 2017. The water temperature (inlet, fixed bed, and outlet), the pH (inlet and outlet), the biogas flow rate, as well as and the CH₄ and the CO₂ contents were continuously recorded with irregular interruptions for maintenance. The flow meter and the gas analyzer proved to be inadequate for continuously and accurately measuring, respectively, the flow and the CH₄ and CO₂ content of biogas produced during the experimental study. For this reason, the response variable used to assess treatment effectiveness was the wastewater calorific value reduction (ΔCV) achieved by passage through the bioreactor. Based on the conservation of energy, the daily total calorific value reduction $CVdt$ (in kJ·day⁻¹) is an indicator of the total amount of pollutant-derived carbon exiting the bioreactor as CO₂ or CH₄ and consequently an indicator of treatment effectiveness. Total solids and other conventional wastewater quality parameters were determined from samples taken on 17 days. The calorific value (CV) of dried total solids was determined over 10 days.

2.3. Surrogate Data

The influent and effluent total solids (respectively TS_i and TS_e in g·L⁻¹) and calorific values (respectively, CV_i and CV_e in kJ·g⁻¹) were used to calculate the sample calorific value reduction (ΔCV) in kJ·L⁻¹ as follows:

$$\Delta CV = (TS_i \times CV_i) - (TS_e \times CV_e) \quad (1)$$

Assuming a positive correlation between the biological activity and temperature, the estimated calorific value reduction ($\widehat{\Delta CV}$) for one day can be calculated as a function of temperature T (in K) and the influent flow rate Q as follows:

$$\widehat{\Delta CV} = a \cdot k_B \cdot \frac{T}{Q} + C \quad (2)$$

where k_B is the Boltzmann constant ($k_B = 1.380649 \times 10^{-23}$ J·K⁻¹), a is a constant in days⁻¹, and C is a constant in J·L⁻¹.

Given the 21 observed or derived data points, we computed a linear regression of Equation (2) (package SciPy optimize curve fit) with $\widehat{\Delta CV}$ expressed in J·L⁻¹. The result

yields a slope $b = 65.5$ and an intercept $C = 178.4$. Then, knowing temperature T and influent flow rate Q , it is possible to estimate $\widehat{\Delta CV}$ with the equation:

$$\widehat{\Delta CV} = 65.5 \cdot \frac{T}{Q} + 178.4 \quad (3)$$

Equation (3) was then used to generate the surrogate data points. The resulting series of 170 data samples of calorific value reduction, including direct and supplemental surrogate data, that constitute a single experiment is referred to as the *reference dataset* and is shown in Figure 1. The data used and a detailed description of the method to generate the reference dataset are presented in a separate document containing supplementary materials. A single experiment is defined by a unique set of four predictors with specific values. This study includes nine different experiments that had different predictors as described below.

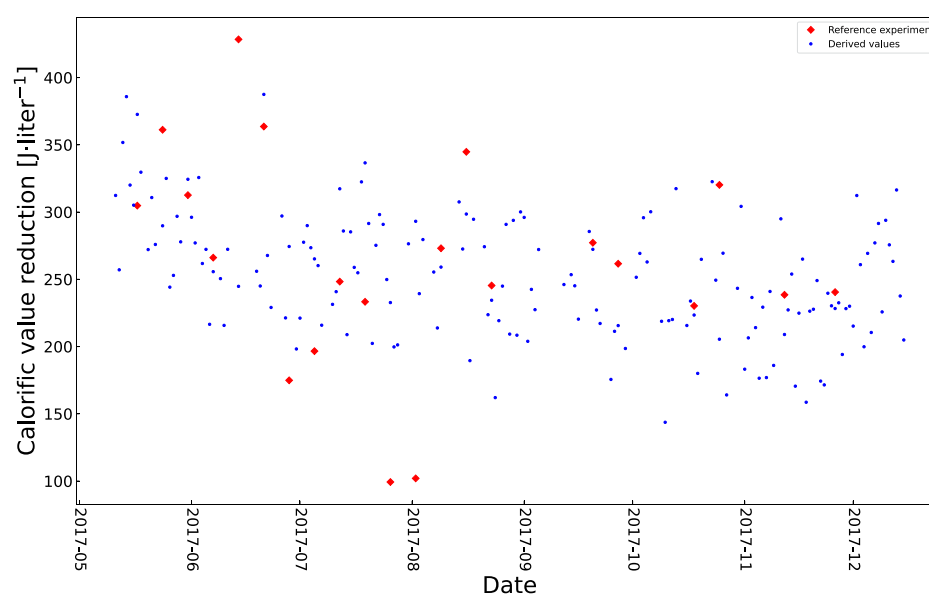


Figure 1. Reference dataset of 170 sampled values of calorific value reduction ΔCV ($J \cdot L^{-1}$). Red diamonds: direct measurements ($n = 21$); blue dots: surrogate values ($n = 170$).

2.4. Mock Experiments

The use of a large and dense dataset is necessary for the training and validation of any ANN model. The experimental design must generate the data points needed to make a valid mapping of the predictor to response variables as assessed by a low value of the error function and the high accuracy of the predictions made using the model. An industrial implementation of the proposed problem-solving approach would require an experimental dataset that fulfills three conditions: (i) the dataset is acquired during operation over a wide range of relevant parameter values; (ii) the dataset is as large and as dense as possible; and (iii) the dataset can be acquired within the technical and economic constraints of the available resources. These requirements could be met using standard probes and data acquisition devices and by manipulating mechanical parameters in a bioreactor test unit designed for this purpose. However, the experimental data described above were acquired before deciding to use ANN models. Consequently, the mechanical set points were not varied, and the size of the experimental dataset is too small to be used to train ANN models. In order to overcome this limitation, additional data derived from mock experiments that use variation of the mechanical set points can be generated for the purpose of demonstrating the approach.

Hence, a mock dataset was created, comprising response values ($CVdt$) derived from the experimental *reference dataset* obtained using the UAF test unit and four predictors: the actual influent flow rate Q and three mechanical factors that were previously described—

i.e., the packing equivalent spherical diameter ESD , the packing material type MAT , and the fixed bed height/diameter ratio HDR . The response values, $CVdt$ of the *mock dataset* were obtained by multiplying the calorific value reduction of the *reference dataset* by a set of coefficients representing the possible effects of the three mechanical factors as shown in Equation (4).

$$CV_{red} = \sum_{i=1}^{170} C_i(Q_i * S_i + Q_i * M_i + Q_i * H_i) \quad (4)$$

where:

CV_{red} = a vector of calorific value reductions for a single mock experiment;

Q_i = the influent flow rate on the i th day;

C_i = the change in the calorific value of the influent stream on the i th day of the reference experiment;

S_i = the spherical diameter effects on calorific value reduction on the i th day;

M_i = the material type effects on calorific value reduction on the i th day;

H_i = the height-to-diameter effects on calorific value reduction on the i th day.

The effects due to a different spherical diameter, material type and height-to-diameter were derived from observations reported in the literature. A detailed description of the method used to generate the mock data and the references on which the ad hoc method are based are available in a separate document containing supplementary materials.

2.5. Experiment Plan to Obtain the Mock Experimental Dataset

The most efficient experimental plan should generate the dataset required to make an accurate ANN model in short duration with few experiments. Orthogonal arrays have the property that in every pair of columns, each of the possible ordered pairs of elements appears the same number of times [34]. The Taguchi arrays are a well-known set of orthogonal experiment designs that have demonstrated their usefulness in making predictive models. The Taguchi L9 experimental plan with four factors set to three different levels requires nine experiments [35]. With reference to the Taguchi L9 layout, a surrogate dataset for use in making an ANN model was generated using only three mechanical factors set to three levels. The fourth factor, the experimentally obtained influent flow rate, was a continuous vector without discrete levels. The influent flow rate comprises all the changes in the input material flux and is the only predictor that changes during an experiment. Consequently, the influent factor levels change in an uncontrolled way. The same influent profile was used in each of the nine mock experiments.

Since each mock experiment had 170 sample points, the total number of response data points was 1530 ($1530 = 170 \times 9$). The total number of predictors (4 factors) was 6120 ($6120 = 1530 \times 4$). Figure 2 shows the reference dataset of daily $CVdt$ (black cross) and the response values from nine separate mock experiments. The density curves on the right side of Figure 2 show that the response values of the mock experiments were distributed over a wide range. The experimental plan with the levels of the mechanical factors and the corresponding $CVdt$ response values (median and mean \pm SD) for each mock experiment are presented in Table 1. This table also shows the rank of the mock experiments. The higher the value of $CVdt$, the better the treatment effectiveness.

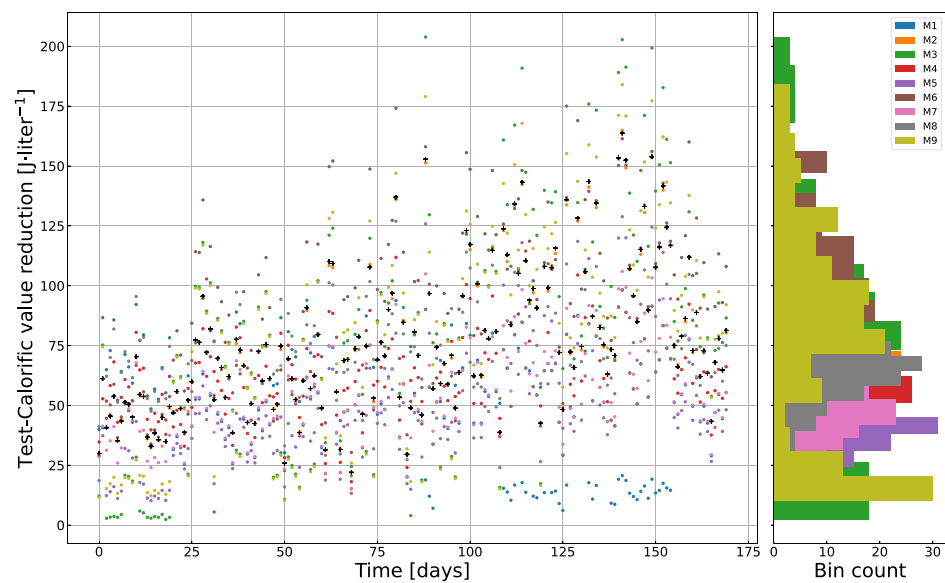


Figure 2. Daily calorific value reduction ($CVdt$). Reference experiment and surrogate data for 1530 sample points from the mock experiment plan. On the righthand side, there are the binned calorific value reductions of the corresponding mock experiments. Black crosses show the reference data points.

Table 1. Experimental plan (mock experiments). Three levels for each mechanical factor. The same influent profile was used in all the mock experiments. *ESD*: equivalent spherical diameter (in mm); *MAT*: packing material type (TWD: torrefied wood chips, PUF: polyurethane foam, PVC: polyvinyl chloride); *HDR*: bed height/diameter ratio. The response data are median (mean \pm SD) $CVdt$ values (daily calorific value reduction) computed for each mock experiment distribution (see Figure 2).

| Experiment Number | ESD | MAT | HDR | CVdt | Rank |
|---------------------------|-----|-----|-----|------------------------|------|
| E1 | 4 | PVC | 0.5 | 48.5 (46.4 \pm 22.9) | 9 |
| E2 (reference experiment) | 12 | TWD | 1.8 | 71.4 (75.9 \pm 30.4) | 4 |
| E3 | 36 | PUF | 3.6 | 79.6 (80.2 \pm 50.4) | 2 |
| E4 | 4 | TWD | 3.6 | 60.9 (61.8 \pm 16.9) | 6 |
| E5 | 12 | PUF | 0.5 | 43.8 (49.7 \pm 27.4) | 8 |
| E6 | 36 | PVC | 1.8 | 92.7 (91.0 \pm 35.5) | 1 |
| E7 | 4 | PUF | 1.8 | 51.3 (54.0 \pm 19.3) | 7 |
| E8 | 12 | PVC | 3.6 | 62.2 (69.1 \pm 18.6) | 5 |
| E9 | 36 | TWD | 0.5 | 82.5 (79.4 \pm 44.7) | 3 |

2.6. Data Preprocessing

The data from the mock experiments were preprocessed to obtain a single dataset for use in training and validating all the models. The Scikit-learn [36], Matplotlib [37], Pandas [38], and Scipy [39] libraries were used for data processing and plotting.

The null hypothesis that the sample calorific value reductions are normally distributed was tested using the Shapiro–Wilk test. The calculated statistic of 0.965 indicates that the sample distribution is not normal (p -value $< 10^{-8}$). The calorific value reductions from the mock experiments are skewed to the right (sample statistic = 0.77). The kurtosis value (0.77, Fisher definition) indicates that the distribution of calorific value reductions has a longer tail than a normal distribution. To obtain a distribution closer to normal, the z-score, log, $\log(1+y)$, and the Yeo–Johnson transformations were compared. The Yeo–Johnson transformation of the response data from the nine mock experiments was found to yield the distribution that is the closest to normal (Shapiro–Wilk statistic = 0.997, p -value = 0.00198). The predictor data were not transformed.

Then, the complete predictor and response value datasets were Min–Max scaled to the range [0, 1]. The prediction accuracy of an MLP model with five hidden layers with 256 units per hidden layer was assessed in terms of RMSE and R^2 .

Compared to training an MLP with raw mock data, training with transformed data improved the accuracy of the predictions. Consequently, the raw data pre-processing during further model development and simulation included the Yeo–Johnson transformation of the response data.

Then, the dataset was split into training and validation sets with a ratio of 80:20, and, respectively, 1224 and 306 sets of predictor and response values corresponding to 1530 sampling events (days). The Kolmogorov–Smirnov statistic (K-S) `ks_2samp` [39] on two samples was computed to test the null hypothesis that the training and validation response datasets were drawn from the same source. The null hypothesis can be rejected at the 99.9% confidence level if the K-S statistic is greater than the critical value (D) of 0.12. The calculated K-S statistic of 0.051 is less than the critical value (p -value = 0.53). Thus, the training and the validation response value datasets have the same distribution and were drawn from the same source.

Depending on the study objectives, the time-series order of the dataset was preserved or shuffled. The `shuffle` parameter was set to `False` when the temporal features of an experiment were preserved and to `True` when the temporal features of an experiment were not included in the dataset. `Random_state` was set to 42 to assure that the preshuffling operation was always the same.

2.7. Predictive Models

The selection of an appropriate predictive model depends on the intended use of the model and on the features of the dataset. In this study, the intended use was to predict a continuous response. Since this is a regression problem, a polynomial regression model was compared to a multilayer perceptron (MLP). Time-series datasets have features that vary over different time periods. In the context of this study, the type of variations include seasonal temperature changes and differences in the flow rates and composition between the day, the night and weekends. Different modeling strategies that include long short-term memory (LSTM) layers, for example, were developed to exploit the specific features of time-series datasets. In contrast, the multilayer perceptron (MLP) architecture is not specifically designed to account for temporal features. Consequently, the model selection approach included an analysis of time dependency, an assessment of conventional regression analysis with a polynomial model, and a comparison to the LSTM and MLP ANN architectures. The LSTM model proved to be less accurate than the MLP model when trained and tested using preshuffled data. Model training with an unshuffled time series dataset is not useful when the aim is to build a model for use in simulating different test cases. Since the aim of this work was to evaluate different predictor values by simulation, the LSTM strategy of response feature modeling was abandoned. The LSTM model and results are presented in the complementary information.

2.7.1. Polynomial Model

A fourth-order regression equation that mapped four predictors to calorific value reduction was fit to unshuffled time series data and then to data that were preshuffled. In both cases, 100 replicate runs were executed using the leave-one-out method to introduce randomness between runs.

The number of regression coefficients was arbitrarily chosen with reference to the four experimentally manipulated predictors. The variance inflation factors (VIFs) were calculated to test for the multicollinearity of the predictor variables. The VIF evaluates the variance of the estimated coefficients and increases to greater than 10 when the predictor variables are highly correlated and values greater than 4 should be investigated [40].

The evaluated array comprised four columns, each representing a predictor variable before Min–Max scaling, and 1530 rows representing the nine experiments ($170 \times 9 =$

1530). The calculated variance inflation factors for ESD, MAT, HDR and influent flow were, respectively: 2.3, 2.0, 2.8, and 3.8.

2.7.2. Multilayer Perceptron (MLP) Model

The multilayer perceptron (MLP) ANN models were built using Python 3.7.6 and the Keras 2.3.0 API running on top of the TensorFlow 2.0 library for machine learning. The sequential model architecture was used. The input object shape was set to 4 ($input_shape = (4,)$). The input object received a single input array containing four vectors corresponding to the four predictors (ESD, MAT, HDR, Q). The output from the input object was passed to the first hidden layer comprised of dense units. The subsequent MLP hidden layers were constructed using identical dense units.

The accuracy of an ANN model can depend on the values of a unique set of weights determined by a single model training and validation run. Since the aim was to develop a general method to build ANN models for use in simulation, the focus was on the model architecture and not on the weights that were obtained from a particular training run. To do this, the initializer seed was set to None so that the unit weights were reinitialized before every model build, train, and test operation.

Different MLP models were built using an iterative process of adding layers and units, setting hyper-parameter values, and evaluating the prediction accuracy. The preliminary best result was a model having 6 hidden layers, each with 192 units per layer. Using this model, different hyper-parameters were then evaluated using semi-automated trials. The RMSprop optimizer [41] was found to perform slightly better than the Adam optimizer. The RMSprop training algorithm maintains a moving average of the square of gradients and divides the gradient by the root of this average. The learning rate was set as low as 1×10^{-5} , which is only 1% of the default value of 1×10^{-3} . The discounting factor for the history/incoming gradient, (ρ) was set as high as 0.99 which is higher than the default value of 0.90. The momentum applied to the optimizer was set as high as 0.9.

The layer hyperparameters `use_bias` and `bias_initializer` were set to True and zeros, respectively. The TruncatedNormal initializer with a mean set to 0.0 and `stddev` set to 0.1 was used to set the initial weights of the hidden layers. The final hyperparameters settings were as follows: `learning rate` = 1×10^{-4} , `rho` = 0.99, `momentum` = 0.05, `epsilon` = 1×10^{-7} , `centered` = True. The learning rate was 10 times less and the momentum slightly more than the default values.

Regularization was applied using dropout. Drop-out layers were placed between the hidden layers to remove a fraction of the preceding layer's units. The rate argument sets a fraction of the drop-out layer units to zero during training. Setting the different percentages [0.1, 1, 2, 3, 4, 5, 10, 25] of the dropout layer units to zero was evaluated. The final model had four dropout layers with the rate argument set to 0.05 (5% set to 0) in each layer.

The model output layer comprised a single dense unit connected to all the units of the preceding hidden layer, and with neither weights, nor bias, nor hyperparameters.

The compiled regression models were trained using a mini-batch gradient descent back-propagation learning algorithm. The default Keras `MeanSquaredError` class instance argument was used to output the average loss of each batch and was also used to evaluate the loss during each training epoch.

I only observed a small improvement in the prediction accuracy when more than 40 samples were used for model training.

Batch size of 1, 2, 4, 8, and 16 (days) were tested. Batch sizes of 1, 2, and 4 days resulted in the same accuracy. Setting the batch size to 8 or 16 days resulted in lower accuracy. I selected a batch size of four for the final model.

To determine the number of hidden layers and hidden layer units, all possible MLP architectures having between 1 and 13 hidden layers and between 8 and 4096 units per hidden layer were built, trained, and evaluated. Model building, training, and evaluation were repeated for each combination of layer and node numbers with the random initialization of the model weights. At the University of Lausanne high-performance computing center,

130 runs were then executed in 4 h using 1 NVIDIA A100 GPU and 10 GB of memory. The prediction accuracy of the validation tests was evaluated in terms of the coefficient of determination (R^2) and the slope of the regression line. There were small differences in the accuracy of the predictions between different model rebuilding operations. The models with more than 512 units per layer were frequently less accurate than the other models. The models with between 32 and 512 units per layer and between 2 and 6 hidden layers yielded similar and acceptable results.

2.8. Use of the Model for Simulation

The study hypothesis that the artificial neural network model can be used to distinguish different sets of up flow anaerobic filter constructive feature set point values by simulation was tested after building the model. The capacity to distinguish different constructive features was evaluated using a *t*-test. The results section contains a complete description of the method.

3. Results

Since all the models were evaluated using the same test data, the most appropriate model can be selected based on the accuracy of the predictions. The accuracy of the predictions was evaluated in terms of the root-mean-square error (RMSE) metric, which is calculated from the differences between true and predicted response values, the coefficient of determination (R^2), and the slope of the regression line of predicted and true values. Both the polynomial and the MLP models were built, trained and validated separately using both the unshuffled time series and preshuffled data sets. The plots of predicted versus true values are shown in Figure 3 (unshuffled) and Figure 4 (preshuffled).

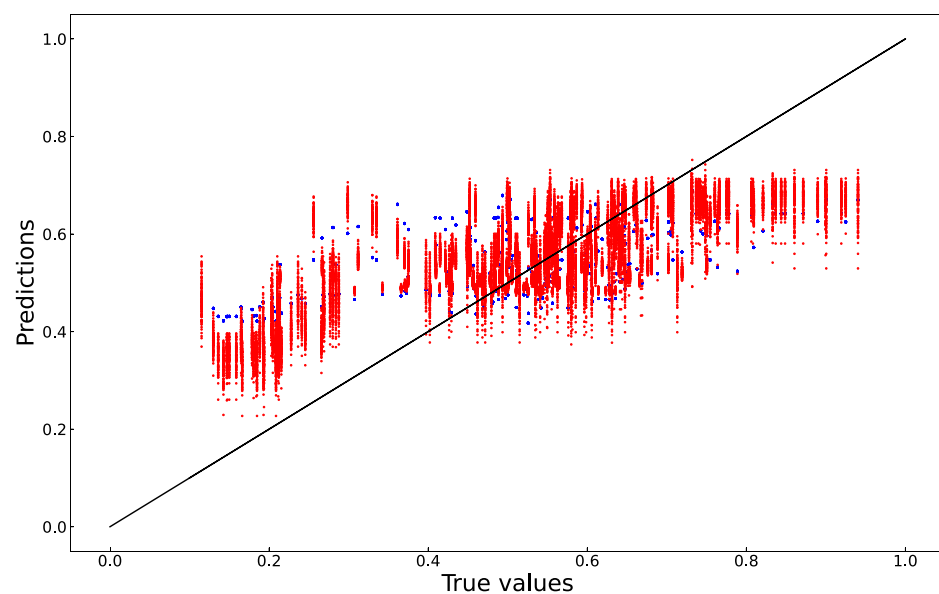
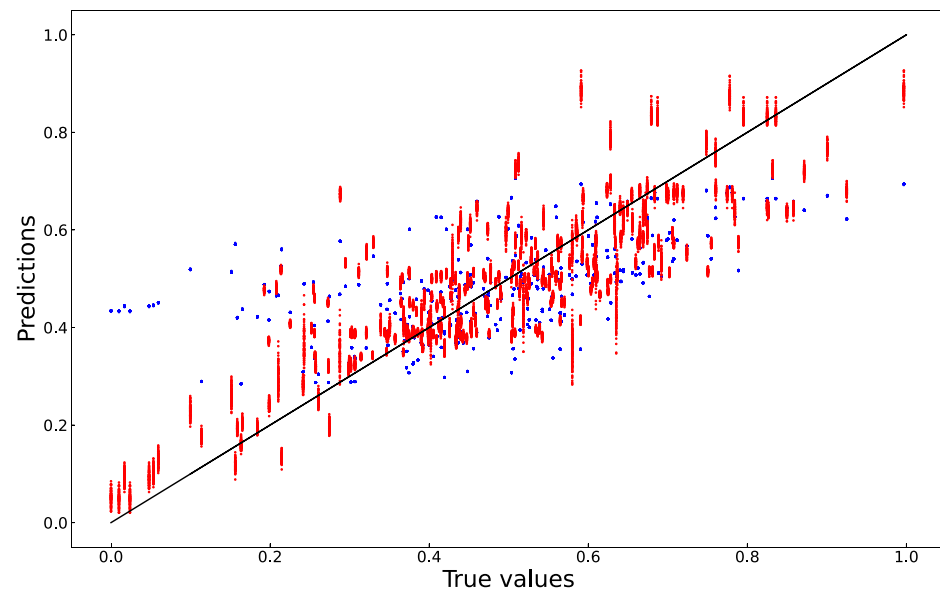


Figure 3. True versus predicted values (unshuffled, true times series data); polynomial model: blue; MLP model: red.

To evaluate the prediction accuracy, 100 runs of the polynomial model were run using the leave-one-out method to introduce randomness between runs. The accuracy of the MLP model was assessed after making 100 training and validation runs with random weight initialization between runs. The results are shown in Table 2. The high slope and the small RMSE demonstrate the accuracy of MLP model. The small variation between runs demonstrates that the prediction accuracy is due to the architecture and not to a set of particular weights and thus contributes to the confidence of the results obtained by simulation.

Table 2. Evaluation of model prediction accuracy.

| Type of Numerical Model | RMSE (%) | R ² | Slope |
|--|----------|----------------|-------|
| Fourth degree polynomial, true time series | 14.9 | 0.29 | 0.24 |
| Fourth degree polynomial, shuffled time series | 14.0 | 0.37 | 0.35 |
| 4-layer MLP, true time series | 12.7 | 0.48 | 0.38 |
| 4-layer MLP, shuffled time series | 10.1 | 0.66 | 0.67 |

**Figure 4.** True versus predicted values (preshuffled data); True versus predicted values (preshuffled data); Polynomial model: blue; MLP model: red.

3.1. Polynomial Model

In terms of RMSE and (R^2), the fourth degree polynomial model was less accurate than the MLP. When unshuffled, true time-series data were used to build the model, and the RMSE was 0.149. The coefficient of determination of a plot the validation test's true and predicted values was 0.29. When preshuffled data were used to build the model, the RMSE was 0.140. The coefficient of determination of a plot the validation test's true and predicted values was 0.37. As shown in Figure 3, the prediction accuracy was very low when the true values were less than 20% and more than 80% of the maximum.

3.2. MLP Model

When unshuffled data were used to build the model, the RMSE was 0.127. The coefficient of determination of a plot of the validation test's true and predicted values was 0.48. When preshuffled data were used to build the model, the RMSE was 0.101. The coefficient of determination of a plot of the validation test's true and predicted values was 0.66. Compared to the polynomial model, the range of accurately predicted values is greater in the MLP model. This difference is also shown in Figure 4 (preshuffled) by the higher slope of the MLP model compared to the polynomial model.

3.3. Time-Series Effects

To investigate the possible time series effects, the autocorrelation function was calculated for 100 days. The autocorrelation coefficient falls to and remains less than 0.5 after an initial 1-day lag and there is a small positive or negative autocorrelation of the time series data after day 16. The calculated Durbin–Watson test statistic (D) was 0.146. The maximum p -value of the correlation coefficient of each time lag, at the 95% level of confidence, was 7.91×10^{-10} . Since the value of the test statistic (D) is less than 2, the time series of the

daily calorific value reduction is positively autocorrelated. These two observations of the response data being autocorrelated imply that the data might have time-dependent features.

However, since the aim of this study was to build a model for use in simulation when fed previously unseen data, the effects of data shuffling and the results of the model validation tests must be considered. The effect of data shuffling prior to training using shuffled data was compared to the effect of training using the actual (not shuffled) time-series data. As shown in Figure 3 (unshuffled), Figure 4 (preshuffled), and in Table 2, the prediction accuracy of the polynomial and MLP models was higher when the models were trained using preshuffled data than when the models were trained using the unshuffled time-series data. The results of model validation tests are shown in Table 2. The predictions made by both the polynomial and the MLP models are less accurate when tested using true time-series data. These results indicate that the performance of the studied processes is not time-series-dependent, and that each day can be treated as an independent sampling point.

3.4. Model Selected for Use in Simulation

In terms of the RMSE, coefficient of determination (R^2), and slope of the regression line, the MLP model yielded a better result than the polynomial model when trained using preshuffled data.

To further compare the polynomial and the MLP models, a t -test was performed to test the null hypothesis that the populations of predictions made using the two different models are the same. The p -values were obtained from 100 runs of the polynomial and the MLP model with random weight initialization between runs. Assuming the equal variances of the two populations and using the Bonferroni correction for 306 samples, the analysis of the p -values showed that the populations are statistically different at the 0.01 alpha level.

The results of the evaluation of 100 replicates of the models in terms of validation mean RMSE, coefficient of determination (R^2), and slope of the regression line are shown in Table 2. Since the four-layer MLP trained using shuffled data was more accurate than the other type of models tested, it was selected for use in simulation.

Figure 5 shows the training and validation losses over 100 epochs. We observed that training and validation loss curves meet after 20 epochs. Although the training and test loss curves separate after 25 epochs, the test loss curve does not rise. The training loss curve decreases at a lower rate before flattening after 80 epochs. The observation that the validation loss does not increase while the training loss remains almost constant indicates that the model is not overfit.

The outcome of model construction, evaluation, and tuning was the selection of the MLP architecture having 264,705 trainable parameters and 4 hidden layers, each with 512 identical dense units. The dataset was preshuffled. The four model input vectors were the packing spherical diameter, packing material type, packed bed height-to-diameter ratio, and the influent flow rate. The only model output was a vector of the daily calorific value reductions (ΔCV expressed in $J \cdot L^{-1}$) of the influent stream pollution load. After training the MLP model to achieve an acceptable loss during validation, the architecture and hyperparameters of the compiled model were saved in HDF5 format for subsequent reloading and use for simulation.

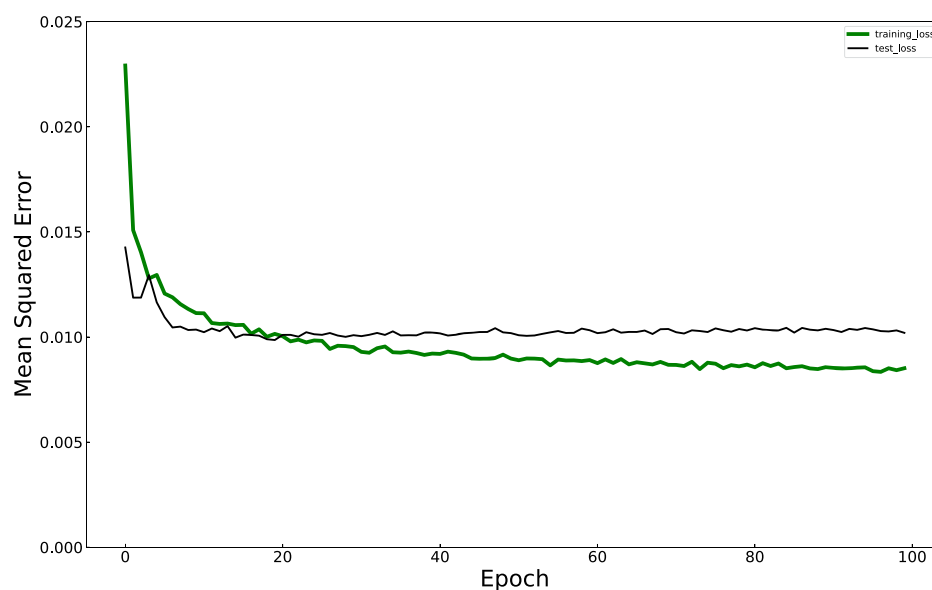


Figure 5. Decrease of MAE during training and validation of the MLP model. Training loss: green; Test loss: black.

3.5. Evaluation of the Experimental Plan

The experimental plan should generate predictions from independent experiments. If the predictors used in the mock experiments were not independent, then some of the experiments would be redundant and consequently bias the experimental results. Redundancy is manifested by the similarity of the response profiles. To investigate a possible redundancy in the experimental plan arrays, response datasets constructed using three different ways of combining two mock experiments were tested: (1) Taguchi-inspired experimental matrix; (2) removing one of the two most similar experiments; and (3) removing one experiment at random. The combination of nine experiments based on the L9 array described above was evaluated using the Kolmogorov–Smirnov test with a significance level set to 0.10 to assess the independence of all combinations of 2 mock experiments from the experimental plan. The corresponding value of the K-S test statistic was 0.18. Out of 36 possible combinations of 2 mock experiments, the calculated K-S statistic was greater than 0.18 in 3 cases. This implies that the experiments included in these three combinations might be redundant. The combination of experiments E3 and E9 had the highest p -value (0.44). Consequently, experiment E3 was removed from the experimental plan. To test the effect of removing the most redundant experiment, the MLP model was then built and evaluated with the experiment E3 removed.

To test the effect of removing a randomly selected experiment before each run, 100 replicates of the MLP model were run using a leave-one-out approach. The prediction accuracy was used to evaluate the effect of removing mock experiments from the experimental plan. Table 3 shows the tested combinations of mock experiments and the resulting RMSE and the coefficient of determination (R^2) of the predictions.

Table 3. Experimental plans evaluated.

| Experiment Combination Description | RMSE | R^2 |
|---|-------|-------|
| Experimental plan as described above (9 experiments) | 0.101 | 0.66 |
| Experimental plan with experiment E3 removed | 0.096 | 0.72 |
| Experimental plan with 1 experiment removed at random | 0.156 | 0.13 |

A slightly better result, in terms of RMSE and the coefficient of determination (R^2), was obtained when redundant experiment E3 was removed. The random removal of a single experiment increased the RMSE and reduced the coefficient of determination (R^2).

The apparent redundancy in the experimental results might be due to real biological or physical causes expressed in the full L9 predictor dataset. For this reason, and because the Taguchi L9 array is a balanced array designed to consider all of the factors and levels equally and to specify a minimal number of experiments, no experiments were removed.

3.6. Use of the MLP for Simulation

The previously trained and saved MLP model was reloaded for use as a *simulation tool* to aid in designing UAF systems for specific use cases. The simulation tool can be used to select the best design specification from comparable simulated test cases. To do this, test cases were defined to explore the possible combinations of influent flow rate, packing material spherical diameter, type of packing material, and fixed bed height-to-diameter ratio within the range of the values used to train the model.

A full factorial design was used to specify a set of 125 permutations of the three mechanical parameters (factors) at five levels. In our simulation step, the high and low-level set-points used were the same as those used in the Taguchi L9 mock experiment design that was followed during the data generation step. The middle level value was the same as the value used during experimental data acquisition. The level values of the mechanical predictors are shown in Table 4.

Table 4. Level values of the mechanical predictors.

| Mechanical Predictors (Factors) | Tested Values (Levels) |
|---------------------------------|------------------------|
| Packing diameter—ESD [mm] | 4, 8, 12, 24, 36 |
| Type of packing material—MAT | 1, 1.85, 2.7, 6.85, 11 |
| Fixed bed H/D ratio—HDR | 0.5, 1.15, 1.8, 2.9, 4 |

For each simulation run, a new daily influent flow rate profile was created *de novo* using the NumPy random uniform generator. All values within the actual minimum and maximum values observed during data acquisition were equally likely to be drawn. To make a comparison between all possible experiments, the same daily influent flow rate profile was used in each experiment conducted during a single simulation run.

As shown in Figure 6, the number of binned influent flow rates observed during the period of data acquisition covered a wide range between 2 and 22 with modes around 200 and 460 L·day⁻¹. Regarding the simulations, the number of binned influent flow rates varied within a more narrow range of between 3 and 14. The coefficient of variation within a bin ranged between 1.3% and 4.8%.

A simulation run consisted in loading the previously built MLP neural network model and then feeding a (21,250, 4) tensor corresponding to 125 experiments of 170 days duration, each having four predictors. Bioreactor performance was evaluated in terms of the total calorific value reduction determined by the simulation. The number of replicate simulations required for a one-sided confidence interval ($\alpha/2$) of 0.5% was estimated using the critical value from the normal distribution and the maximum sample standard deviation [42]. The minimum number of replicates is seven. The sum of the daily total predicted Min–Max scaled calorific value reduction was calculated for 20 replicates of the simulated experiments. The results of the 125 simulated experiments were then ranked according to the mean calorific value reduction.

The box plot (Figure 7) shows the means and the distribution of values within replications of a single experiment and the range of predicted cumulative total calorific value reductions. Identical influent streams were fed to all 125 experiments included in a single replication. However, each replication had a unique randomly generated influent stream flowrate profile.

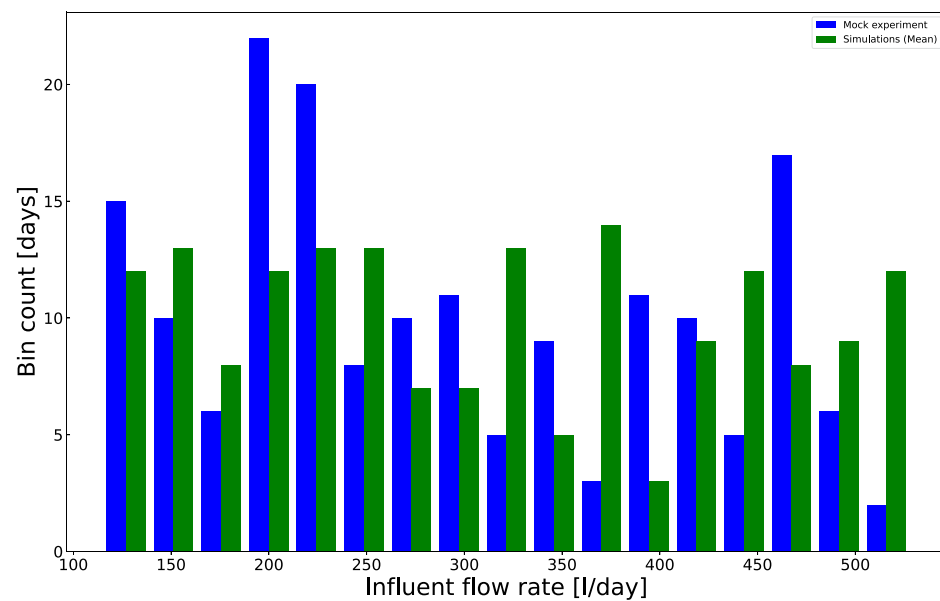


Figure 6. Influent flow rate: experiments (blue bars); simulations (green bars).

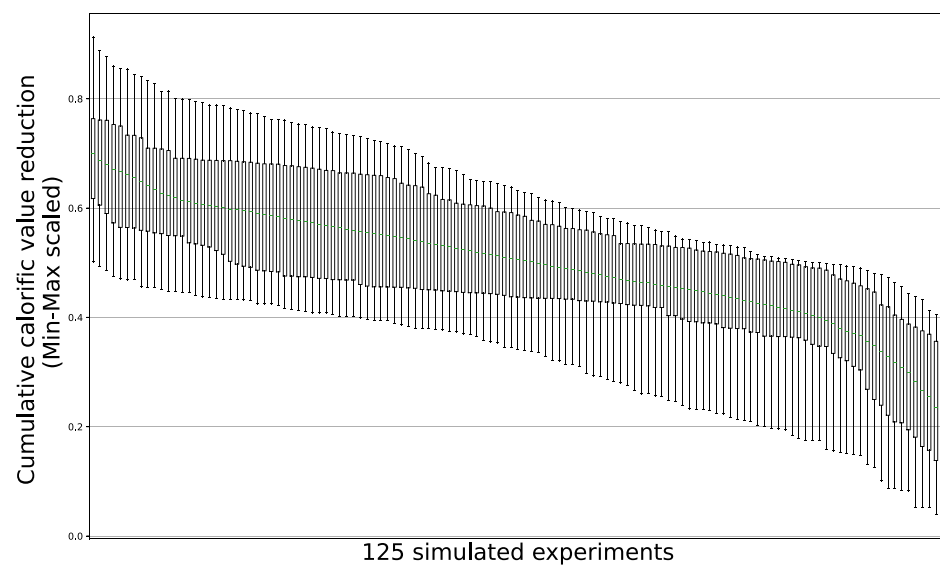


Figure 7. Box plot of the predicted calorific value reduction. Results of 125 simulated experiments using the MLP model. The boxes enclose all values in the 25th through the 75th percentiles. The whisker reach is set to two times the range of boxes. The green lines show the mean of the simulation runs.

The null hypothesis of no difference between the means of the three highest ranking experiments was tested using a *t*-test and assuming equal variances in the sample populations. Simulated experiments SE125, SE119 and SE100 yielded the highest total calorific value reduction. Since the simulated experiment SE119 had the second highest ranking mean total calorific value reduction, the significance of the difference between SE125 and SE119 was tested. The calculated *p*-value of the difference between experiments SE125 and SE119 is 0.0108, indicating that the rejection of the null hypothesis is expected to be erroneous in much less than 2% of experiments. The calculated *p*-value of the difference between experiments SE125 and SE100 is 0.0027, indicating that the rejection of the null hypothesis is expected to be erroneous in less than 1% of experiments. Thus, the null hypothesis is rejected and the mechanical set point values used in experiment SE125 are expected to yield the highest total calorific value reduction during more than 98% of the tested influent flow regimes.

The spherical diameter was 36 in the 3 highest ranked experiments. In the three highest ranked experiments, the material type code was 1 or 1.85, indicating that the packing material with a smooth surface is appropriate. The height-to-diameter ratio was 4 in the 2 highest ranking experiments and 2.9 in the third highest ranking experiment. Consequently, the bioreactor should be long.

The same MLP model was reloaded and used to simulate influent flow rates in the low range of 127–321 L per day. The simulation results show that using the mechanical parameters $ESD = 4$, $MAT = 11$, and $HDR = 4$ results in the highest calorific reduction. Consequently, if the influent flow rate is low, then a small diameter and rough surfaced packing material should be used.

These design recommendations are based on mock data and are meant to only illustrate the method. The design recommendations are not meant for use in constructing a real upflow anaerobic filter. The design recommendation for use in constructing a real upflow anaerobic filter must be based on real measurements made during experiments.

4. Discussion

The method and workflow presented in this study are intended to improve the entire environmental bioprocesses development workflow from raw data collection to design specification. The present study demonstrates, for the first time, the workflow and use of an MLP as a simulation tool to specify constructive features of an upflow anaerobic filter. The contributions of this study are compared to existing work and the advantages of the proposed method are discussed below.

4.1. Comparison to Existing Methods

Many laboratory and industrial studies describe the use of numeric models for wastewater treatment system identification and process performance forecasting based on water quality parameters. For example, in the context of a study to optimize the operating cost of a large activated sludge wastewater treatment plant, Koncsos [43] built an MLP neural network model to predict the concentrations of components such as ammonium ($\text{NH}_4\text{-N}$) and nitrate ($\text{NO}_3\text{-N}$). The ANN model was trained and validated using a surrogate dataset generated using the variants of established activated sludge models (ASMs). For $\text{NH}_4\text{-N}$ and $\text{NO}_3\text{-N}$, the correlation coefficients (R^2) of the predictions made by the ANN with ASM dataset were, respectively, 0.889 and 0.954. For suspended solids, expressed as COD, the correlation coefficients (R^2) of the ANN with ASM dataset were 0.987 for the 0–5 min interval and 0.843 for the 10–20 min future interval. In contrast, the intended use of the model built during the present study was simulation, and not operational process forecasting. Moreover, the purpose of simulation was to design equipment and not to optimize process control. The predictive model built during the present study did not include influent water quality parameters. The constructive features of the bioreactor and the influent flow rate were the only predictors of process performance. Additionally, the prediction interval was 170 days and the process performance was assessed by the cumulative calorific value reduction during this period rather than parameter prediction during a short interval. Forecasting models were also applied to anaerobic processes. For example, in a study of industrial starch processing wastewater treatment by an upflow anaerobic sludge blanket (UASB) bioreactor, Antwi et al. [44] built an MLP to predict COD removal from influent (COD, NH_4 , pH, OLR) and effluent parameters (VFA and biogas yield) in the predictor dataset. The coefficient of determination (R^2) of the data reserved for testing was 0.87. This work is an accurate mapping of the selected prediction to response variables that helps to understand a complex process. In comparison, the present study did not include effluent parameters in the predictor dataset and the aim was to develop a simulation tool. With the goal of intensifying biogas production in a modern wastewater treatment plant, Sakiewicz et al. [45] built general regression neural network (GRNN), radial basis function (RBF) network and MLP models of data acquired during 3 years of continuous plant operation. This study focused on identifying the important

parameters that affect process performance. Their analysis approach made it possible to conclude that the technical control process parameters have a more important influence on process performance than the wastewater quality indicators. Consequently, attention should be focused on mechanical features such as process control settings rather than on the wastewater stream quality and microbiological features of the anaerobic digestion process. Importantly, this result suggests that it might be possible to build a high-performance numerical model of technical environmental bioprocesses without the knowledge of the conventional water quality parameters and the biochemical parameters used in the IWA mathematical modeling approach. In contrast to the present study, they did not consider mechanical constructive features and they did not run simulations using neural network models.

Although less frequently used than operational parameters, some mechanical parameters have been included in numeric models of wastewater treatment processes. For example, the packing material dimension, packing material surface features, the bioreactor geometry and the influent flow rate are known to have an effect on UAF performance. Bolte et al. [46] included the flow rate, surface area and porosity of the support medium in a mathematical model used to simulate anaerobic filters treating dilute waste streams. In an expired patent [47], it is claimed that the diameter of the biomass support spheres must be sufficient to generate the surface shear forces required to shed microorganisms, and thereby prevent bed clogging. In a laboratory study of an upflow anaerobic packed bed reactor treating organic waste streams, Tay et al. [48] examined the significance of media factors such as media roughness, specific surface area, porosity and pore size on treatment performance. They found that media pore size and porosity play significant roles in the performance of the reactor. Different media types can have different pore sizes and porosity. For example, PVC has a smooth, non-porous surface, and foam can have a rough, porous surface. Consequently, different packing material diameters, material types, sizes and surface characteristics were investigated as predictors in the present study. Tests of anaerobic digestion of food industry wastewater in 10-liter anaerobic filters fed batches of influent from an industrial site showed that most of the organic substrates were used in the bottom of the reactor [49]. This implies that there is an optimal fixed bed length. If the fixed bed is too short, then the bioreactor is less effective. If the fixed bed is too long, then the bioreactor construction costs and the space requirements are unnecessarily high. These observations led to the decision to investigate the fixed bed height-to-diameter ratio during the present study. By integrating these important, and previously recognized, mechanical features into a single ANN predictive model, the present study contributes to the UAF design process.

Additionally, this study contributes to the development of methods to build and use neural networks in the field of environmental systems. In the present study, the intended use of the artificial neural network for simulation is very different from other possible uses such as fault detection or real-time forecasting. This use requires a predictive model that fits a large dataset acquired during a long time period. The suitability for use in the field of wastewater treatment, and in particular to model the upflow anaerobic filter, was an additional requirement. Considering these requirements, the polynomial, MLP, and LSTM models are among the natural candidates. Referring again to the study of Sakiewicz et al. [45], general regression neural network (GRNN), and the radial basis function (RBF) network and MLP models of data acquired during 3 years of continuous plant operation were built and evaluated. They determined that the MLP model with two hidden layers with 13 and 9 units per layer was the most accurate. In a laboratory study that compared the predictive ability of the numerical models of upflow anaerobic filters, Yilmaz [50] demonstrated that an MLP having only one hidden layer with six units more accurately predicted the methane yield than the conventional multi-linear regression or the Stover–Kincannon models. Using a surrogate dataset obtained from the simulations of an activated sludge wastewater treatment plant made using the ASM model, Demir [51] built an MLP neural network model to predict COD, phosphorus and nitrogen removal

from the influent concentrations and operating parameters that control internal recycle. The coefficient of determination (R^2) of true to predicted values was between 0.95 and 0.99. This study showed how the design engineer can use the model to determine design and operational parameters by simulation and sensitivity analysis. Instead of using sensitivity analysis, we selected the design parameters by ranking the results of simulations made using all possible permutations of three design parameters. In the context of the forecast modeling of N_2O concentrations in wastewater treatment plants, Hwangbo et al. [52] observed that an LSTM model outperformed a DNN model in terms of forecast prediction accuracy. They noted that the LSTM-based forecasting modeling depends on the response data, whereas all sensor data can be used as predictors in DNN models. Consequently, they concluded that the LSTM-based model is not suitable for process modeling that includes every measured variable. This fundamental difference between the LSTM and MLP model architectures might explain the inaccuracy, in terms of (R^2), of the predictions made during the present study using the LSTM model. In agreement with this conclusion, the MLP model was selected for use in simulation during the present study. The accuracy and capacity of a numeric model for a specific use case, such as simulation, partly depends on the model architecture. Since there is no established methodology to determine the architecture, the subject is a popular research topic. Dibaba et al. [53] used a differential evolution (DE) approach to tune an ANN model of an upflow anaerobic contactor (UAC) for biogas production from vinasse. The ANN model predicted effluent chemical oxygen demand (COD), total suspended solids (TSSs), volatile fatty acids (VFAs) and biogas production from the characteristics of the influent vinasse. Their DE method involved the random manipulation of the MLP dimensions and hyperparameters and then selection of the best configuration based on the error. Using the DE method, the performance of the ANN model, in terms of the coefficient of determination (R^2), improved from 0.891 to 0.921. In comparison, the present study included a broad search and the evaluation of 130 different combinations of the number of hidden layers and units in order to select the optimal MLP architecture. A semi-automated trial and error approach to determine the best MLP hyperparameter settings in terms of the coefficient of determination (R^2). Building on this experience, MLP models having from 1 to 13 hidden layers and from 8 to 4096 units per layer. The time requirement to build and train a large neural network is no longer an issue. The selected neural network had 4 hidden layers, each having 256 units. This model was built and trained for 100 epochs in 3 h using a laptop (HP EliteBook 840 G4, Intel Core i5-7200U CPU@2.50 GHz, 8 Go RAM).

Forecasting a parameter value requires a model that maps predictor values, that change in time and often in an irregular way, to a response value. In contrast, a numeric model that simulates fixed mechanical values does not have to account for changes in the predictors. A model that simulates mechanical values is comparable to a physical prototype with features that closely resemble those of the final machinery. Prototype testing consists of executing an experimental plan to expose the physical device to the expected service conditions. Given the benefit of the close resemblance of the prototype to the final machinery, an overparameterized neural network might be desirable. The MLP used in this study was highly overparameterized with many more trainable model parameters than data points. Contrary to conventional practice, recent evidence has shown that there are significant benefits to having more parameters than data points when training a neural network [54]. The authors cite a state-of-the-art image recognition system, with a parameter-to-data point ratio of 400, as an example of such large-scale overparameterization. Early stopping during training results in an overparameterized neural network that is effective for interpolation. Consequently, the model used for simulation was highly overparameterized and trained until the number of epochs was sufficient to assure a low prediction error. The ratio of model parameters to data points was 35.

Although the present study did not aim to develop a soft sensor, the calorific value reduction could be considered as a composite parameter that describes process performance. ANN-based soft sensors are used to predict unmeasured or impossible to measure variables.

For example, Pisa et al. [55] developed an MLP-based soft sensor to predict effluent total nitrogen since this parameter cannot be measured online. The prediction structure included LSTM cells that predicted effluent concentrations 4 h in the future based on input data from the previous 10 h. It is currently not possible to measure effluent calorific reduction online. The MLP predictive model developed during this study could be improved to build an ANN-based soft sensor for calorific value reduction.

4.2. Contribution of the Proposed Method

The conventional workflow to develop environmental bioprocesses could be improved, for example, by reducing the time and resource requirements for laboratory experimentation and by collecting data under real operating conditions. To meet this requirement, a data acquisition system was designed to make it possible to easily vary real mechanical features and to measure process parameters under real, industrial, operating conditions. This study demonstrated that it is not necessary to include conventional bioprocesses parameters such as kinetic constants, microbial inhibition factors, and hydrodynamic parameters that are difficult to measure in industrial conditions. For this reason, the proposed method can be described as a “black-box” approach.

Since the prediction accuracy of an ANN model depends on the number of raw data points, the data acquisition unit and the experimental plan should make it possible to acquire a very large dataset and to vary the relevant mechanical parameters. The most practical way to address this problem is to allow these parameters to naturally vary during the experiment thereby including the complexity of the industrial conditions in the dataset used to build the ANN. In the present study, the influent flow rate widely and irregularly varied during experimental data acquisition. The proposed method also takes advantage of advances in sensing and data acquisition techniques that make it possible to measure process parameters that might be meaningful when using the “black-box” approach but meaningless in a conventional dynamic chemical equilibrium approach to bioprocess modeling. Thus, the present study contributes to the movement towards the use of in-line measurement and data intensive methods during process development.

This study demonstrated that it is possible to acquire the data required to build a useful numeric model without laboratory experiments and that a balanced experimental array design can help to improve the prediction accuracy of the model. The experimental plan should reflect the highly variable quality of the influent wastewater stream in terms of composition, flow rate and temperature that, because of economic constraints, cannot be regulated in an industrial context. The use of a modified Taguchi experimental design to obtain a balanced array was demonstrated. However, any experimental design that leads to the development of a high-quality model can be used. The quality of any numerical model can be assessed by examining how closely model implementation agrees with the underlying physics of the studied subject, the prediction error as defined by the relative difference between the predicted value and the true measured value, and by the accuracy of the predictions made from data that represent a wide range of operating conditions. In the proposed approach, the underlying physics was accounted for by easily measured parameters such as exit gas and liquid flow rate and mechanical constructive features.

This study demonstrated the use of an ANN model of experimentally acquired data as a simulation tool to explore bioreactor performance under new conditions that were not experimentally tested. The present study demonstrates, for the first time, the workflow and use of a MLP to specify the mechanical constructive features of an upflow anaerobic filter. Using the proposed approach is expected to save time. The duration of the mock experimental phase was only 5.5 months, whereas the simulations covered more than 58 years of continuous operation.

The method and workflow demonstrated by this study are intended to be generally applicable to the design of the environmental process equipment and constructed natural systems.

4.3. Future Implementation

In future work, the proposed method could be implemented using three bioreactor test units operated simultaneously under real industrial conditions. The values of mechanical parameters would be set to cover the high and low extremes according to an experimental plan designed so that in every pair of columns, each of the possible ordered pairs of elements appears the same number of times as is characteristic of orthogonal arrays. The plan would also include the vectors of continuously varying process operational parameters such as flow rate or a water quality parameters that do not conform to the rules of orthogonal arrays. The experimental plan would be implemented to gather as many data points as possible during a short duration. For example, the influent flow rate and methane production measurements could be made continuously during 2 months. The resulting dataset would be used to build an MLP model for use in predicting the bioreactor performance under any set of simulation conditions lying within the range of values observed during the experiment. Thus, the model would make it possible to explore different equipment dimensions and operating set-point values that were not experimentally tested.

It is also possible to use the proposed method to model existing datasets. The methods of surrogate data generation described in this study and dataset reconfiguration could be used to create mock experiments.

5. Conclusions

This study demonstrated the development of a multilayer perceptron (MLP) artificial neural network (ANN) predictive model and its use to determine bioreactor mechanical design parameters by simulation. The main findings of this study were:

- The predictions made using the MLP were more accurate than those made using the polynomial model (coefficients of determination (R^2), respectively, of 0.66 and 0.37);
- The MLP model used for simulation should be defined by architecture and not by a particular set of hidden-layer unit weight initializations;
- The reloaded MLP model can be used for simulation using previously unseen input vectors;
- The differences in the predictions are significant (p -value < 0.02, range = mean \pm 20% of the full range);
- The ranked order of simulation results can be used to select the set of mechanical specifications that will result in the highest equipment performance;
- A practical on-site data collection plan can be used to reduce the resource requirements for new process development.

The demonstration that different sets of mechanical design parameters yield significantly different predicted performance confirms the capacity of the model to distinguish differences in equipment design parameters. By making it possible to explore a wide range of mechanical parameter set points without costly and time-consuming laboratory experiments, the implementation of the method is expected to lead to improved, more efficient, design specifications of environmental bioprocessing equipment.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/su14137959/s1>. References [49,56–62] are cited in the supplementary materials.

Funding: This study was funded by the Swiss Academy of Engineering Sciences (SATW), grant number 2020-001 <https://www.satw.ch/en/> (accessed on 24 May 2022). M.M. was partially supported by Innosuisse—Swiss Innovation Agency and SCCER BIOSWEET <https://www.innosuisse.ch/inno/en/home.html> (accessed on 24 May 2022).

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: Alessandro E.P. Villa contributed to the data analysis and neural network model methodology development.

Conflicts of Interest: The author declares no conflict of interest.

References

- IWA. *International Water Association (IWA)*; IWA: London, UK, 2020.
- Dochain, D.; Vanrolleghem, P.A. *Dynamical Modelling & Estimation in Wastewater Treatment Processes*; IWA Publishing: London, UK, 2001; p. 9781780403045. <https://doi.org/10.2166/9781780403045>.
- Barton, P.I.; Lee, C.K. Modeling, simulation, sensitivity analysis, and optimization of hybrid systems. *ACM Trans. Model. Comput. Simul.* **2002**, *12*, 256–289. <https://doi.org/10.1145/643120.643122>.
- Dochain, D. (Ed.) *Bioprocess Control*; ISTE: London, UK, 2008; 248p.
- Ramaswamy, S.; Cutright, T.; Qammar, H. Control of a continuous bioreactor using model predictive control. *Process Biochem.* **2005**, *40*, 2763–2770. <https://doi.org/10.1016/j.procbio.2004.12.019>.
- Lima, F.V.; Rawlings, J.B. Nonlinear stochastic modeling to improve state estimation in process monitoring and control. *AIChE J.* **2011**, *57*, 996–1007. <https://doi.org/10.1002/aic.12308>.
- Liotta, F.; Chatellier, P.; Esposito, G.; Fabbicino, M.; hullebusch, E.D.V.; Lens, P.N.L.; Pirozzi, F. Current Views on Hydrodynamic Models of Nonideal Flow Anaerobic Reactors. *Crit. Rev. Environ. Sci. Technol.* **2015**, *45*, 2175–2207. <https://doi.org/10.1080/10643389.2015.1010426>.
- Teixeira, A.P.; Alves, C.; Alves, P.M.; Carrondo, M.J.T.; Oliveira, R. Hybrid elementary flux analysis/nonparametric modeling: Application for bioprocess control. *BMC Bioinform.* **2007**, *8*, 30. <https://doi.org/10.1186/1471-2105-8-30>.
- Zhang, Q.; Hu, J.; Lee, D.J.; Chang, Y.; Lee, Y.J. Sludge treatment: Current research trends. *Bioresour. Technol.* **2017**, *243*, 1159–1172. <https://doi.org/10.1016/j.biortech.2017.07.070>.
- Haimi, H.; Mulas, M.; Corona, F.; Vahala, R. Data-derived soft-sensors for biological wastewater treatment plants: An overview. *Environ. Model. Softw.* **2013**, *47*, 88–107. <https://doi.org/10.1016/j.envsoft.2013.05.009>.
- Fisher, O.J.; Watson, N.J.; Porcu, L.; Bacon, D.; Rigley, M.; Gomes, R.L. Multiple target data-driven models to enable sustainable process manufacturing: An industrial bioprocess case study. *J. Clean. Prod.* **2021**, *296*, 126242. <https://doi.org/10.1016/j.jclepro.2021.126242>.
- Tetko, I.; Villa, A. An Enhancement of Generalization Ability in Cascade Correlation Algorithm by Avoidance of Overfitting/Overtraining Problem. *Neural Process Lett.* **1997**, *6*, 43–50. <https://doi.org/10.1023/A:1009610808553>.
- Tetko, I.V.; Aksenova, T.I.; Volkovich, V.V.; Kasheva, T.N.; Filipov, D.V.; Welsh, W.J.; Livingstone, D.J.; Villa, A.E. Polynomial neural network for linear and non-linear model selection in quantitative-structure activity relationship studies on the internet. *SAR QSAR Environ. Res.* **2000**, *11*, 263–280. <https://doi.org/10.1080/10629360008033235>.
- Tetko, I.V.; Tanchuk, V.Y.; Kasheva, T.N.; Villa, A.E.P. Estimation of aqueous solubility of chemical compounds using E-state indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488–1493. <https://doi.org/10.1021/ci000392t>.
- Harada, L.H.P.; da Costa, A.C.; Maciel Filho, R. Hybrid neural modeling of bioprocesses using functional link networks. *Appl. Biochem. Biotechnol.* **2002**, *98–100*, 1009–1023. <https://doi.org/10.1385/abab:98-100:1-9:1009>.
- Ławryńczuk, M. Modelling and nonlinear predictive control of a yeast fermentation biochemical reactor using neural networks. *Chem. Eng. J.* **2008**, *145*, 290–307. <https://doi.org/10.1016/j.cej.2008.08.005>.
- Rene, E.R.; López, M.E.; Kim, J.H.; Park, H.S. Back propagation neural network model for predicting the performance of immobilized cell biofilters handling gas-phase hydrogen sulphide and ammonia. *BioMed Res. Int.* **2013**, *2013*, 463401. <https://doi.org/10.1155/2013/463401>.
- Sharghi, E.; Nourani, V.; Ashrafi, A.; Gökçekuş, H. Monitoring effluent quality of wastewater treatment plant by clustering based artificial neural network method. *Pol. J. Environ. Stud.* **2019**, *164*, 86–97. <https://doi.org/10.5004/dwt.2019.24385>.
- Delnavaz, M.; Farahbakhsh, J.; Talaiekhosani, A.; Mehdinezhad Nouri, K. Predicting Removal Efficiency of Formaldehyde from Synthetic Contaminated Air in Biotrickling Filter Using Artificial Neural Network Modeling. *J. Environ. Eng.* **2019**, *145*, 04019056. [https://doi.org/10.1061/\(asce\)ee.1943-7870.0001566](https://doi.org/10.1061/(asce)ee.1943-7870.0001566).
- de Menezes, P.L.; de Azevedo, C.A.V.; Eying, E.; Neto, J.D.; de Lima, V.L.A. Artificial neural network model for simulation of water distribution in sprinkle irrigation. *Rev. Bras. Eng. Agrícola E Ambient.* **2015**, *19*, 817–822. <https://doi.org/10.1590/1807-1929/agriambi.v19n9p817-822>.
- Venkatesh Prabhu, M.; Karthikeyan, R. Comparative studies on modelling and optimization of hydrodynamic parameters on inverse fluidized bed reactor using ANN-GA and RSM. *Alex. Eng. J.* **2018**, *57*, 3019–3032. <https://doi.org/10.1016/j.aej.2018.05.002>.
- Salehi, E.; Askari, M.; Aliee, M.H.; Goodarzi, M.; Mohammadi, M. Data-based modeling and optimization of a hybrid column-adsorption/depth-filtration process using a combined intelligent approach. *J. Clean. Prod.* **2019**, *236*, 117664. <https://doi.org/10.1016/j.jclepro.2019.117664>.
- Arismendy, L.; Cárdenas, C.; Gómez, D.; Maturana, A.; Mejía, R.; Quintero, M.C.G. Intelligent System for the Predictive Analysis of an Industrial Wastewater Treatment Process. *Sustainability* **2020**, *12*, 6348. <https://doi.org/10.3390/su12166348>.
- Witherow, J.L.; Coulter, J.B.; Ettinger, M.B. Anaerobic Contact Process for Treatment of Suburban Sewage. *J. Sanit. Eng. Div.* **1958**, *88*, 1–13.

25. Richard, R.; Dague, R.E.M.; Pfeffer, J.T. Anaerobic Activated Sludge. *J. Water Pollut. Control Fed.* **1966**, *38*, 220–226.
26. Young, J.C.; McCarty, P.L. The anaerobic filter for waste treatment. *J. Water Pollut. Control Fed.* **1969**, *41*, 160.
27. Genung, R.K.; Million, D.L.; Hancher, C.W.; Pitt, W.W., Jr. *Pilot Plant Demonstration of an Anaerobic Fixed-Film Bioreactor for Wastewater Treatment*; Oak Ridge National Lab.: Oak Ridge, TN, USA, 1978.
28. Manariotis, I.D.; Grigoropoulos, S.G. Municipal-Wastewater Treatment Using Upflow-Anaerobic Filters. *Water Environ. Res.* **2006**, *78*, 233–242.
29. Córdoba, P.R.; Riera, F.S.; Sineriz, F. Temperature effects on upflow anaerobic filter performance. *Environ. Technol. Lett.* **1988**, *9*, 769–774. <https://doi.org/10.1080/09593338809384631>.
30. Young, J.C. Factors Affecting the Design and Performance of Upflow Anaerobic Filters. *Water Sci. Technol.* **1991**, *24*, 133–155. <https://doi.org/10.2166/wst.1991.0222>.
31. Bhattacharya, J.; Dev, S.; Das, B. Design of Wastewater Bioremediation Plant and Systems. In *Low Cost Wastewater Bioremediation Technology*; Butterworth-Heinemann: Oxford, UK, 2018; pp. 265–313.
32. Zhu, G.Y.; Zamamiri, A.; Henson, M.A.; Hjortsø, M.A. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chem. Eng. Sci.* **2000**, *55*, 6155–6167. [https://doi.org/10.1016/S0009-2509\(00\)00208-6](https://doi.org/10.1016/S0009-2509(00)00208-6).
33. Tarjányi-Szikora, S.; Oláh, J.; Makó, M.; Palkó, G.; Barkács, K.; Zárny, G. Comparison of different granular solids as biofilm carriers. *Microchem. J.* **2013**, *107*, 101–107. <https://doi.org/10.1016/j.microc.2012.05.027>.
34. Kacker, R.N.; Lagergren, E.S.; Filliben, J.J. Taguchi's Orthogonal Arrays Are Classical Designs of Experiments. *J. Res. Natl. Inst. Stand. Technol.* **1991**, *96*, 577–591. <https://doi.org/10.6028/jres.096.034>.
35. Roy, R.K. *A Primer on the Taguchi Method*, 2nd ed.; Society of Manufacturing Engineers: Dearborn, MI, USA, 2010.
36. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
37. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. <https://doi.org/10.1109/MCSE.2007.55>.
38. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; van der Walt, S., Millman, J., Eds.; SciPy: Austin, TX, USA, 2010; pp. 56–61. <https://doi.org/10.25080/majora-92bf1922-00a>.
39. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.
40. Iain Pardeoe, L.S.; Young, D. STAT 462 Applied Regression Analysis, 2021. Available online: <https://online.stat.psu.edu/stat462/node/180/> (accessed on 24 May 2022).
41. Hinton, G.; Srivastava, N.; Swersky, K. Neural Networks for Machine Learning Lecture 6a Overview of mini-Batch gradient descent. *Cited* **2012**, *14*, 2.
42. NIST. NIST/SEMATECH e-Handbook of Statistical Methods, Section 7.2.2.2. 2021. Available online: <https://www.itl.nist.gov/div898/handbook/prc/section2/prc222.htm> (accessed on 14 March 2022).
43. Koncsos. Bioreactor Simulation with Quadratic Neural Network Model Approximations and Cost Optimization with Markov Decision Process. *Period. Polytech. Civ. Eng.* **2020**, *64*, 614–622. <https://doi.org/10.3311/ppci.14734>.
44. Antwi, P.; Li, J.; Meng, J.; Deng, K.; Quashie, F.K.; Li, J.; Boadi, P.O. Feedforward neural network model estimating pollutant removal process within mesophilic upflow anaerobic sludge blanket bioreactor treating industrial starch processing wastewater. *Bioresour. Technol.* **2018**, *257*, 102–112. <https://doi.org/10.1016/j.biortech.2018.02.071>.
45. Sakiewicz, P.; Piotrowski, K.; Ober, J.; Karwot, J. Innovative artificial neural network approach for integrated biogas—Wastewater treatment system modelling: Effect of plant operating parameters on process intensification. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109784. <https://doi.org/10.1016/j.rser.2020.109784>.
46. Bolte, J.P.; Nordstedt, R.A.; Thomas, M.V. Mathematical Simulation of Upflow Anaerobic Fixed Bed Reactors. *Trans. ASAE* **1984**, *27*, 1483–1490. <https://doi.org/10.13031/2013.32991>.
47. Trösch, W.; Kiefer, W.; Lohmann, K.; Dürold, H. Porous Inorganic Support Spheres Which Can Be Cleaned of Surface Biomass under Fluidized Bed Conditions. U.S. Patent 4,987,068, 22 January 1991.
48. Joo-Hwa, T.; Kuan-Yeow, S.; Jeyaseelan, S. Media Factors Affecting the Performance of Upflow Anaerobic Packed-Bed Reactors. *Environ. Monit. Assess.* **1997**, *44*, 249–261. <https://doi.org/10.1023/a:1005740708480>.
49. Berardino, S.D.; Costa, S.; Converti, A. Semi-continuous anaerobic digestion of a food industry wastewater in an anaerobic filter. *Bioresour. Technol.* **2000**, *71*, 261–266. [https://doi.org/10.1016/S0960-8524\(99\)00080-2](https://doi.org/10.1016/S0960-8524(99)00080-2).
50. Yilmaz, T. Modeling the performance of upflow anaerobic filter (UAF) reactors treating paper-mill wastewater using neural networks. *Sci. Res. Essays* **2013**, *8*, 1896–1905.
51. Demir, S. Artificial Neural Network Simulation of Advanced Biological Wastewater Treatment Plant Performance. *Sigma J. Eng. Nat. Sci.* **2020**, *38*, 1713–1728.
52. Hwangbo, S.; Al, R.; Chen, X.; Sin, G. Integrated Model for Understanding N₂O Emissions from Wastewater Treatment Plants: A Deep Learning Approach. *Environ. Sci. Technol.* **2021**, *55*, 2143–2151.
53. Dibaba, O.R.; Lahiri, S.K.; T'Jonck, S.; Dutta, A. Experimental and Artificial Neural Network Modeling of a Upflow Anaerobic Contactor (UAC) for Biogas Production from Vinasse. *Int. J. Chem. React. Eng.* **2016**, *14*, 1241–1254. <https://doi.org/10.1515/ijcre-2016-0025>.

54. Thomson, N.C.; Greenwald, K.; Lee, K.; Manso, G.F. The Computational Limits of Deep Learning. *arXiv* **2020**, arXiv:2007.05558.
55. Pisa, I.; Santín, I.; Vicario, J.; Morell, A.; Vilanova, R. ANN-Based Soft Sensor to Predict Effluent Violations in Wastewater Treatment Plants. *Sensors* **2019**, *19*, 1280. <https://doi.org/10.3390/s19061280>.
56. Araujo, P.; Astray, G.; Ferrerio-Lage, J. A.; Mejuto, J. C.; Rodriguez-Suarez, J. A.; Soto, B. Multilayer perceptron neural network for flow prediction. *J. Environ. Monit.* **2011**, *13*, 35–41.
57. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780.
58. Smith, L. C.; Elliot, D. J.; James, A. Mixing in upflow anaerobic filters and its influence on performance and scale-up. *Water Res.* **1996**, *30*, 3061–3073.
59. Olah, C. Understanding LSTMs, 2015. Available online: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed on 14 March 2022).
60. Escudié, R.; Conte, T.; Steyer, J.P.; Delgenès, J.P. Hydrodynamic and biokinetic models of an anaerobic fixed-bed reactor. *Process Biochem.* **2005**, *40*, 2311–2323.
61. Seabold, S.; Perktold, J. Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010.
62. Thirumurti, D.; Effects of mixing velocity on anaerobic fixed film reactors. *Water Res.* **1988**, *22*, 517–523.

An Artificial Neural Network for Simulation of an Upflow Anaerobic Filter Wastewater Treatment Process

Complementary information

Mark McCormick

University of Lausanne
Faculty of Business and Economics (HEC Lausanne)
Department of Information Systems (DESI)
Neuroheuristic research group

April 14, 2022

1 Summary

The complementary information includes the data and the methods used to prepare the datasets. Descriptive statistics describing the raw and the transformed reference data are shown. The process of generating mock data is described in detail. The polynomial, multi-layer perceptron and long short-term memory models are described in detail. Additional results that support method development and the conclusions are reported.

2 Material and Methods

Table 1 shows 21 sets of values of influent and effluent total solids (respectively TS_i and TS_e in g.L^{-1}) and calorific values (respectively CV_i and CV_e in kJ.g^{-1}). The calorific values of 11 data points that were derived from measured total solids are shown in bold text. The calorific values of the remaining 10 data points were derived from the measured TS_i and TS_e . These values were used to calculate the sample calorific value reduction (ΔCV) in kJ.L^{-1} as follows:

$$\Delta CV = (TS_i \times CV_i) - (TS_e \times CV_e) \quad (1)$$

In addition to ΔCV , the fixed bed temperature T (in $^{\circ}\text{C}$) and the influent flow rate Q (in L.day^{-1}), are also indicated in Table 1.

As described in the article text, these data were used to make the regression equation.

$$\widehat{\Delta CV} = a \cdot k_B \cdot \frac{T}{Q} + C \quad (2)$$

where k_B is the Boltzmann constant ($k_B = 1.380649 \times 10^{-23} \text{ J.K}^{-1}$), a is a constant in days^{-1} , and C is a constant in J.L^{-1} . This equation can also be rewritten as:

$$\widehat{\Delta CV} = b \cdot \frac{T}{Q} + C \quad (3)$$

Given the 21 points observed, we compute a linear regression of Equation 3 (package SciPy optimize curve fit) with $\widehat{\Delta CV}$ expressed in J.L^{-1}

The result yields a slope $b = 65.48$ and an intercept $C = 178.39$. Then, knowing temperature T and influent flow rate Q it is possible to estimate $\widehat{\Delta CV}$ with the equation:

$$\widehat{\Delta CV} = 65.48 \cdot \frac{T}{Q} + 178.39 \quad (4)$$

A noise vector was generated to add supposed effects that Equation 4 does not account for. The noise vector was obtained by adding vectors obtained from the mean (μ) and standard deviation (σ) of the ratio of measured temperature T and influent flow rate Q according to Equation 5 :

$$noise = 0.5 \cdot \mu \cdot \frac{T}{Q} + 1.2 \cdot \sigma \quad (5)$$

The noise vector was then added to the vector obtained from taking the ratio of the measured temperature T and influent flow rate Q . Figure 2 shows the noise vector. Equation 6 was then used to generate the surrogate data points.

$$\widehat{\Delta CV} = 65.48 \cdot \left(\frac{T}{Q} + noise \right) + 178.39 \quad (6)$$

The regression equation was used to generate surrogate data from measured temperature and influent flow rates. Thus, the resulting series of 170 data samples of calorific value reduction — referred to as the *reference dataset* — included direct and supplemental surrogate data.

Table 1: Daily average measurements derived from the experimental UAF bioreactor of influent (TS_i) and effluent (TS_e) total solids, influent (CV_i) and effluent (CV_e) calorific values, fixed bed temperature (T) and influent flow rate (Q). The daily calorific value reduction (ΔCV) was derived following Equation 1.

| Day | TS_i (kg.L ⁻¹) | CV_i (kJ.kg ⁻¹) | TS_e (kg.L ⁻¹) | CV_e (kJ.kg ⁻¹) | ΔCV (kJ.L ⁻¹) | T (°C) | Q (L.day ⁻¹) |
|-------------|------------------------------|-------------------------------|------------------------------|-------------------------------|-----------------------------------|----------|----------------------------|
| 17-MAY-2017 | 0.07 | 5.88 | 0.06 | 1.78 | 0.3048 | 22.6 | 142.6 |
| 24-MAY-2017 | 0.09 | 5.34 | 0.06 | 1.99 | 0.3612 | 24.0 | 151.2 |
| 31-MAY-2017 | 0.08 | 5.25 | 0.06 | 1.79 | 0.3126 | 22.4 | 144.0 |
| 07-JUN-2017 | 0.06 | 5.92 | 0.05 | 1.90 | 0.2661 | 19.2 | 196.4 |
| 14-JUN-2017 | 0.08 | 7.41 | 0.06 | 2.74 | 0.4284 | 29.1 | 223.2 |
| 21-JUN-2017 | 0.08 | 5.62 | 0.05 | 1.72 | 0.3636 | 35.9 | 144.0 |
| 28-JUN-2017 | 0.05 | 5.10 | 0.03 | 2.67 | 0.1749 | 23.9 | 220.2 |
| 05-JUL-2017 | 0.05 | 5.55 | 0.05 | 1.84 | 0.1966 | 27.4 | 216.0 |
| 12-JUL-2017 | 0.06 | 5.68 | 0.05 | 1.85 | 0.2483 | 23.1 | 194.4 |
| 19-JUL-2017 | 0.06 | 5.68 | 0.05 | 2.15 | 0.2333 | 30.4 | 165.6 |
| 26-JUL-2017 | 0.03 | 6.06 | 0.05 | 1.65 | 0.0993 | 23.1 | 316.8 |
| 02-AUG-2017 | 0.03 | 5.88 | 0.04 | 1.86 | 0.1020 | 32.1 | 309.6 |
| 09-AUG-2017 | 0.06 | 5.86 | 0.05 | 1.57 | 0.2731 | 23.1 | 288.0 |
| 16-AUG-2017 | 0.07 | 5.68 | 0.06 | 0.88 | 0.3448 | 23.9 | 312.6 |
| 23-AUG-2017 | 0.06 | 5.64 | 0.05 | 1.86 | 0.2454 | 22.3 | 367.2 |
| 20-SEP-2017 | 0.06 | 5.76 | 0.04 | 1.71 | 0.2772 | 23.1 | 230.2 |
| 27-SEP-2017 | 0.08 | 4.20 | 0.06 | 1.24 | 0.2616 | 20.6 | 385.2 |
| 18-OCT-2017 | 0.05 | 5.62 | 0.04 | 1.69 | 0.2303 | 15.8 | 408.0 |
| 25-OCT-2017 | 0.07 | 6.10 | 0.05 | 2.36 | 0.3202 | 14.3 | 309.6 |
| 12-NOV-2017 | 0.07 | 6.50 | 0.06 | 3.50 | 0.2385 | 14.6 | 316.0 |
| 26-NOV-2017 | 0.08 | 6.50 | 0.07 | 4.30 | 0.2405 | 12.0 | 448.0 |

As shown in Figure 1, the CV reduction increased with the ratio of influent temperature to flow.

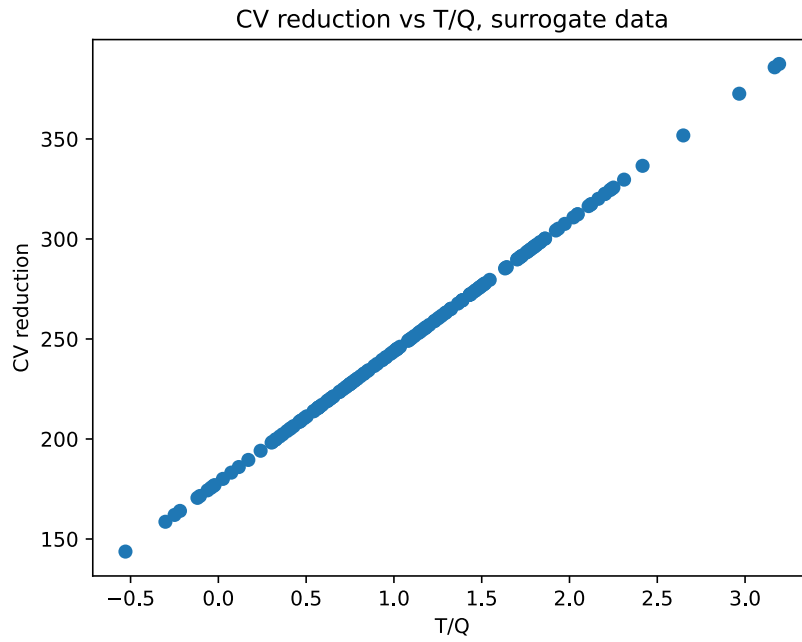


Figure 1: Ratio of CV reduction versus temperature/influent flow

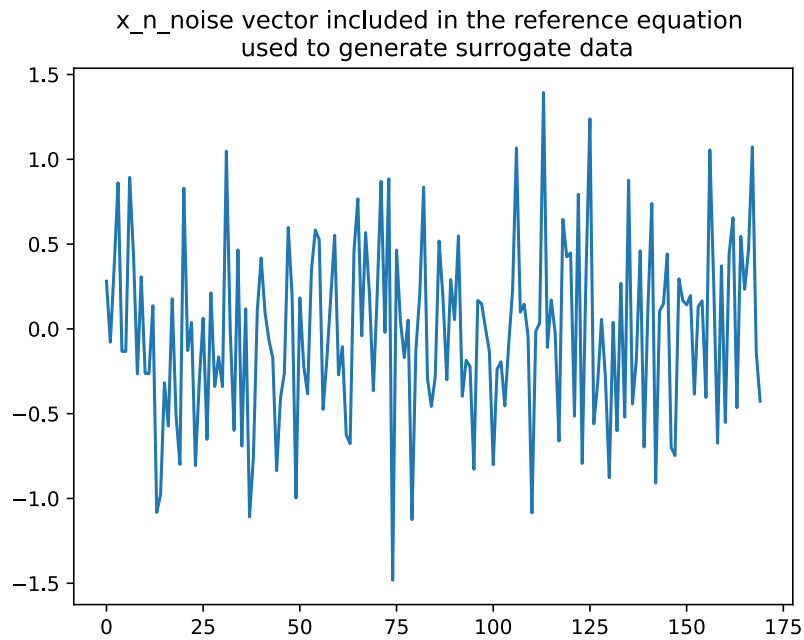


Figure 2: Noise vector included in the regression equation used to generate surrogate data.

Mock data generation

2.0.1 Packing equivalent spherical diameter (ESD)

The selected mock packing equivalent spherical diameters $ESD = [4, 12, 36]$ with the middle level the same as the the ESD of the experimental bioreactor and the other values 3 times smaller than and 3 times larger than the experimental value. Larger packaging material leads to an increase in void space between individual pieces of packing material. An increase in the effectiveness of the treatment, expressed in terms of biogas production, results from better mixing that is achieved when the void space volume increases [4]. To include the effect of large void space, the coefficient increases exponentially with the influent flow rate when the value of ESD is 36 mm. In a study of a bioreactor containing 3.5 m tubular packing, the efficiency of the biological reactions was limited by substrate transfer that is very sensitive to substrate loading and methane flow rate [6]. It is also known that inadequate mixing of the sludge bed occurs at very low flow rates, and that very high flow rates lead to shear and biomass loss [8, 6]. In the dataset from the mock experiments, these effects were created using coefficients that included the ratio of the influent flow rate to the maximum influent flow rate and by setting a threshold value below which effectiveness is reduced due to very poor mixing.

2.0.2 Packing material type (MAT)

Based on the relative measured activity of biofilms on samples of polyvinyl chloride (PVC) chips, torrefied wood chips (TWD), and polyurethane foam (PUF) taken from the experimental bioreactor, the coefficients for the factor MAT were set to respectively $c_{MAT(PVC)} = 1.0$, $c_{MAT(TWD)} = 2.7$, and $c_{MAT(PUF)} = 11.0$. Since biogas production improves mixing [4], a coefficient increased effectiveness when the material type was polyurethane foam. Polyurethane foam packing effectiveness is reduced at low flow rates due to biomass clogging the interstitial space. PVC packing effectiveness is reduced at high flow rates due to biomass shear loss. Consequently during mock experiments the value of $c_{MAT(PVC)}$ was further reduced when the influent flow rate was greater than 80% of the maximum.

2.0.3 Fixed bed height to diameter ratio (HDR)

Since the optimal HDR was not known, three levels — $HDR = [0.5, 1.8, 4.0]$ — were selected arbitrarily with the middle level the same as the the HDR of the experimental bioreactor. It has been reported that most of the organic substances are consumed in the bottom part of the bioreactor [2]. This implies that the bed is too long when the organic load is low and that a well mixed long bed has the highest capacity for CV reduction (ΔCV). Consequently, the value of the coefficient was made to increase with the flow rate to an upper limit that was proportional to the bioreactor height to diameter ratio.

As described in the article, the predictors were set according to an experimental plan that was based on the Taguchi L9 design with the 4th factor (predictor) set to the experimental influent flow rate.

The 4 predictors were:

- Equivalent spherical diameter (ESD)
- Material type (MAT)
- Height to diameter ratio (HDR)
- Influent flow rate (Q_{in})

Figure 3 below shows the values of the predictors used in the 9 mock experiments in sequence.

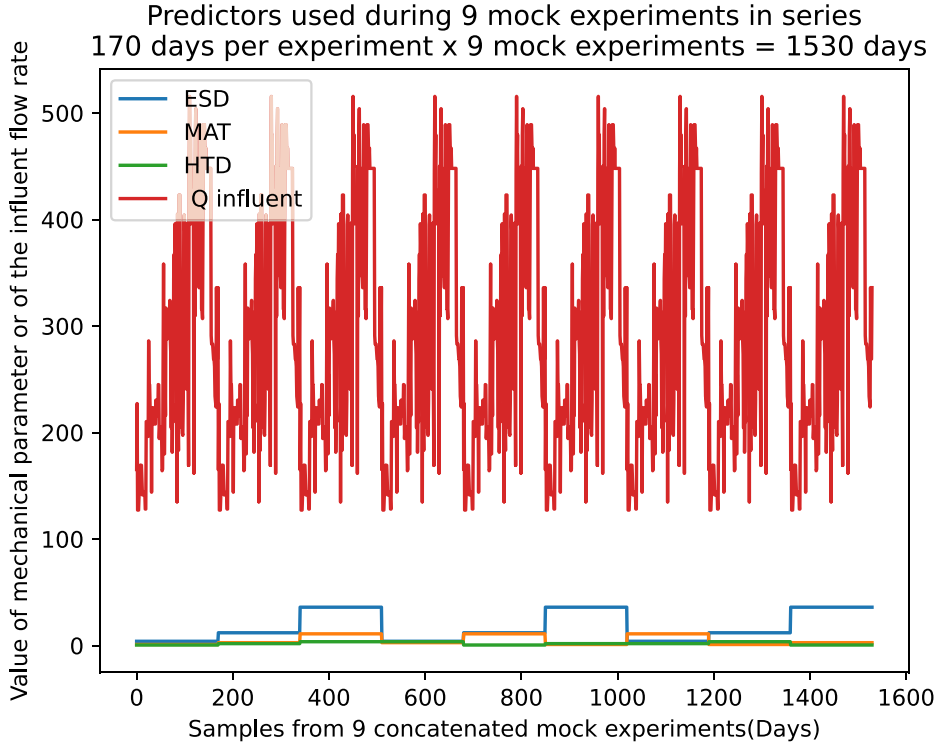


Figure 3: **Value of the predictors. 9 mock experiments in sequence**

Predictors: Sphere diameter (ESD), Material type (MAT), Height to diameter ratio (HDR), Influent flow rate (Q influent).

An *ad hoc* function was used to generate the calorific value reduction from the sum of the ESD, MAT and HDR components. The daily calorific value reduction was obtained from the product of the influent flow rate and the predictor effects. The *ad hoc* function is shown in the article and below.

$$CV_{red} = \sum_{i=1}^{170} C_i (S_i * Q_i + M_i * Q_i + H_i * Q_i)$$

where,

CV_{red} = a vector of calorific value reductions for a single mock experiment

Q_i = the influent flow rate on the i th day

C_i = the reference change in the calorific value of the influent stream on the i th day

S_i = the spherical diameter effects on calorific value reduction on the i th day

M_i = the material type effects on calorific value reduction on the i th day

H_i = the height to diameter effects on calorific value reduction on the i th day

The following Python code was used to generate mock data using the *ad hoc* function.


```

S=[] # ESD component of total CV reduction
M=[] # MAT component of total CV reduction
H=[] # HDR component of total CV reduction
Qmax = np.max(IFR_ref_all) #516 l/day Maximum influent flow rate
i=0
h=0
SF = 1550 # Scale factor to adjust the value predicted by each of the 3 features.
SC = 1 # sphere diameter adjustment factor
SC4 = 0.4 # Threshold below which the CVred increases exponentially in proportion to flow rate when Q
is below SC4*Qmax and SD = 4 because small SD is well adapted to low flow rates.
# Predictors. 9 experiments with values from the range described in the MDPI article text.
exper_lit = [(4.0, 1.0, 0.5), (12.0, 2.7, 1.8), (36.0, 11.0, 3.6), (4.0, 2.7, 3.6), (12.0, 11.0, 0.5), (36.0, 1.0, 1.8),
(4.0, 11.0, 1.8), (12.0, 1.0, 3.6), (36.0, 2.7, 0.5)]
for h in range(0,len(exper_lit)): # Mock experiments kJl x l/day = kJ/day
for i in range(len(IFR_1)):

    S.append(np.multiply((CV_ref_kjl[i]),( # aprox 0 - 0,5 kJl
1+
np.where(exper_lit[h][0]==12.0, SF*0.33*IFR_1[i]/Qmax, 0)+
np.where(exper_lit[h][0]==36.0 and IFR_1[i] > 0.8*Qmax , SF*0.1, 0)+
np.where(exper_lit[h][0]==36.0 and IFR_1[i] > SC4*Qmax , 0.8*SF*0.33*(1-(IFR_1[i]/Qmax)), 0)+
np.where(exper_lit[h][0]==36.0 and IFR_1[i] < SC4*Qmax , 0.1*SF*0.33*np.exp(-IFR_1[i]/Qmax), 0) +
np.where(exper_lit[h][0]==36.0 and IFR_1[i] < 0.45*Qmax , -SF*0.05, 0) +
np.where(exper_lit[h][0]==4.0 and IFR_1[i] < SC4*Qmax , 0.4*SF*0.33*np.exp(-IFR_1[i]/Qmax), 0)+
np.where(exper_lit[h][0]==4.0 and IFR_1[i] < 0.8*Qmax , 2*(SF*4/12)*0.33*IFR_1[i]/Qmax, 0)+
np.where(exper_lit[h][0]==4.0 and IFR_1[i] > 0.8*Qmax , 0.01*SF*0.33*(1-np.exp(IFR_1[i]/Qmax)),0)+
np.where(exper_lit[h][0]==4.0 and IFR_1[i] < 0.45*Qmax , -SF*0.01, 0)+
0)*SC)*1)

    M.append(np.multiply((CV_ref_kjl[i]),(
1+
np.where(exper_lit[h][1]==2.7, SF*0.33*IFR_1[i]/Qmax, 0) +
np.where(exper_lit[h][1]==11.0 and IFR_1[i]> 0.8*Qmax, SF*0.2, 0)+
np.where(exper_lit[h][1]==11.0 and IFR_1[i]> 0.3*Qmax, 0.3*SF*0.33*np.sin((3.14/2)*(IFR_1[i]/Qmax)), 0)
+
np.where(exper_lit[h][1]==11.0 and IFR_1[i]< 0.3*Qmax, -0.1*SF*0.33*np.exp(-IFR_1[i]/Qmax), 0)+
np.where(exper_lit[h][1]==11.0 and IFR_1[i]< 0.3*Qmax, -0.05*SF*0.33, 0)+
np.where(exper_lit[h][1]==1.0 and IFR_1[i]< SC4*Qmax, 0.5*SF*0.33*np.exp(-IFR_1[i]/Qmax), 0)+
np.where(exper_lit[h][1]==1.0 and IFR_1[i]< 0.8*Qmax, 11*(SF*1/11)*0.33*IFR_1[i]/Qmax, 0)+
np.where(exper_lit[h][1]==1.0 and IFR_1[i]> 0.8*Qmax, 0.03*SF*0.33*(-np.exp(-IFR_1[i]/Qmax)),0)+
0)*SC)*1)

    H.append(np.multiply((CV_ref_kjl[i]), (
1+
np.where(exper_lit[h][2]==1.8, SF*0.33*IFR_1[i]/Qmax, 0) +
np.where(exper_lit[h][2] == 0.5 and IFR_3[i]<Qmax*0.5/3.6, SF*0.33*IFR_3[i]/Qmax, 0)+
np.where(exper_lit[h][2] == 0.5 and IFR_3[i]>0.5/3.6, 0.4*SF*0.33*IFR_3[i]/Qmax, 0)+
np.where(exper_lit[h][2] == 3.6 and IFR_3[i]< Qmax*1.8/3.6, SF*0.33*IFR_3[i]/Qmax, 0)+
np.where(exper_lit[h][2] == 3.6 and IFR_3[i]>Qmax*1.8/3.6, 0.8*SF*0.33*IFR_3[i]/Qmax, 0) +
0)*SC)*1)

IFR_Mtest = np.concatenate((np.asarray(S), np.asarray(M), np.asarray(H)), axis=1).T

```

The contribution of each predictor and the total CV reduction on each day are shown separately in the

figure 4 below.

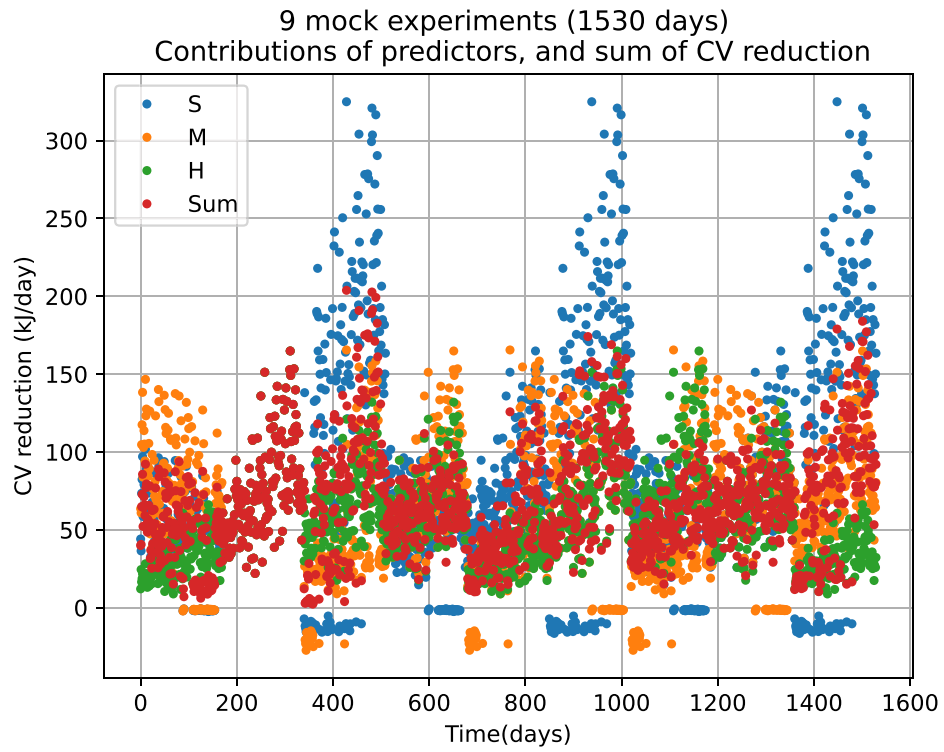


Figure 4: **Calorific value reduction (ΔCV)**. Components: Sphere diameter (S), Material type (M), Height to diameter ratio (H), Sum of the component contributions (Sum)

The histogram below (Figure 5) shows the range of calorific value reductions and the difference between the 9 mock experiments.

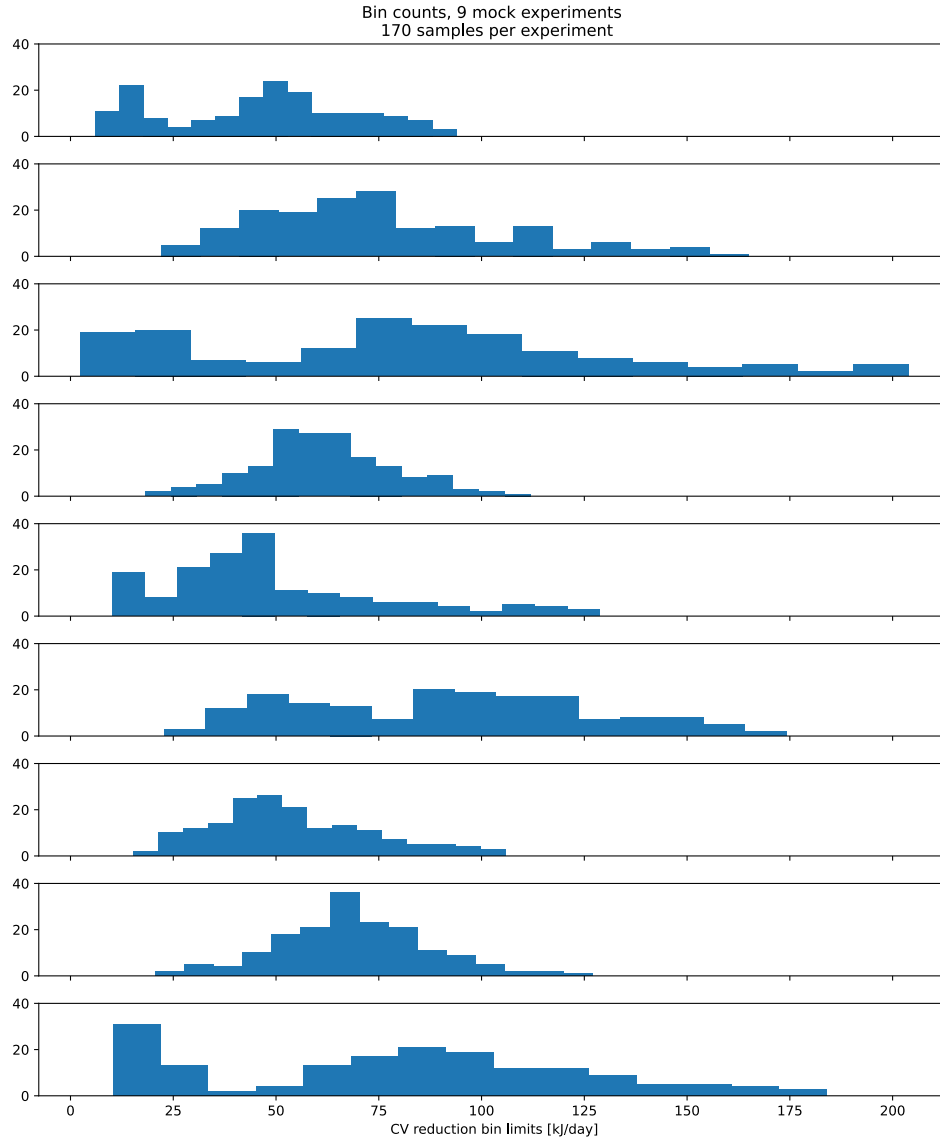


Figure 5: **Binned calorific value reduction (ΔCV)**. 9 experiments, 170 sample values per experiment. Experiment numbers 1 - 9 (from top to bottom)

For the purpose of comparison, response data was generated from the predictors used in the experiment using the mock response data generation function. The predictors were: ESD=12, MAT=2.7, HDR=1.8

with influent flow the same as the experiment. Figure 6 shows a plot of experimentally acquired data verses mock data.

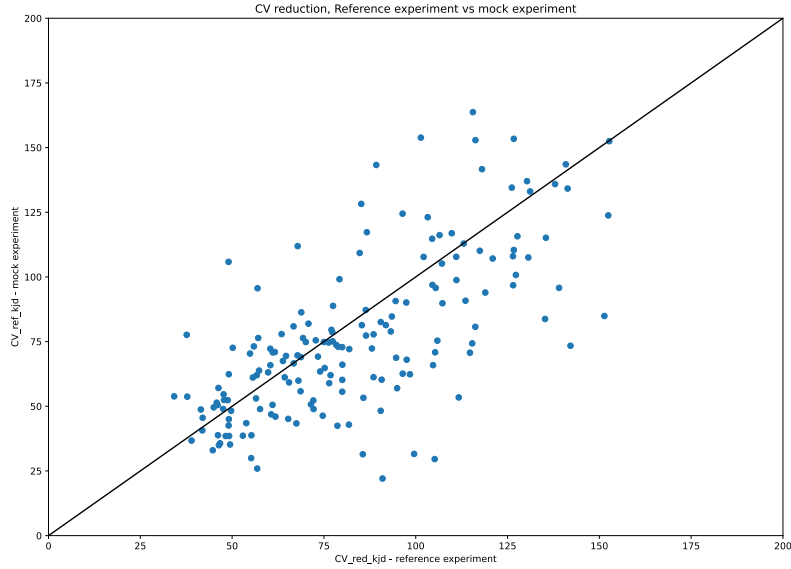


Figure 6: **CV reduction, reference experiment versus data generated using the mock response data generation function.**

2.0.4 Transformation of the response data

Different transformations of the response data were evaluated to obtain a more normally shaped distribution of the response data. The Yeo-Johnson transformation was found to yield the distribution that was most like a normal distribution. The histograms of the raw and the transformed response data are shown in Figure 7. The cumulative CVreduction for the different transformations is shown in Figure 8.

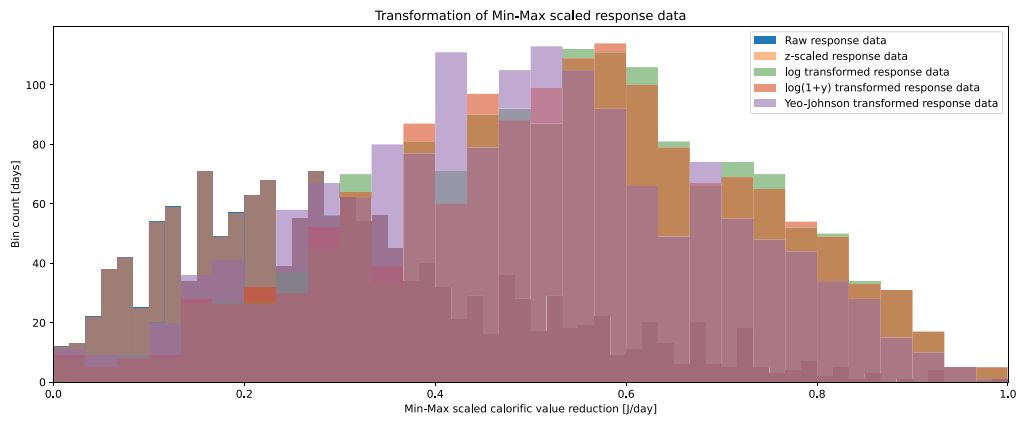


Figure 7: **Mock data transformation** Distribution of calorific value reductions 9 mock experiments, 1530 data points

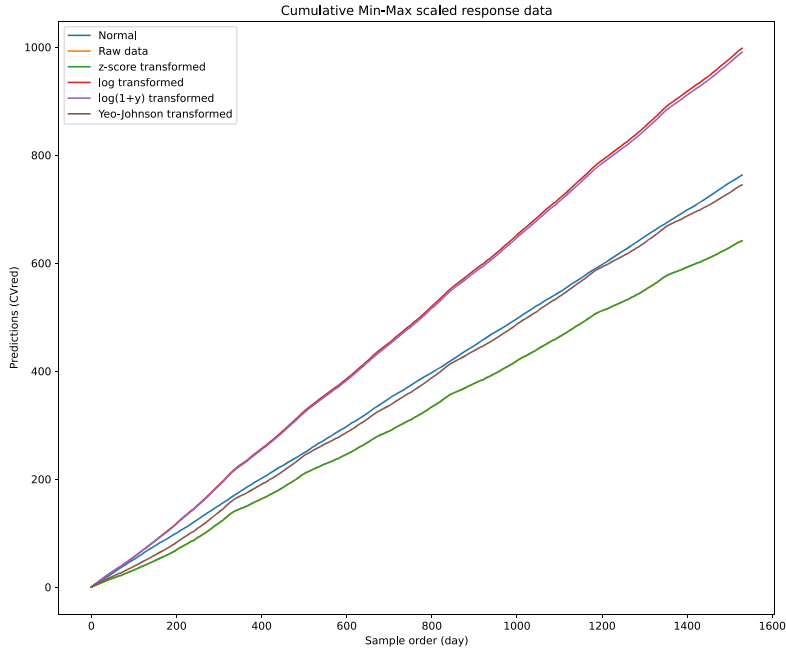


Figure 8: **Mock data transformation** Cumulative CV reduction, 9 mock experiments, 1530 data points

2.0.5 Analysis of time dependency

Autocorrelation evaluates the similarity between current and lagged data points in a time series dataset. The Durbin-Watson test was performed to determine if there is autocorrelation in the series of daily calorific value reduction. The null hypothesis is no autocorrelation. The value of the calculated test statistic lies between 0 and 4. A value of 2 indicates no autocorrelation. Values below 2 indicate positive autocorrelation. Values above 2 indicate negative autocorrelation. The calculated D-W test statistic was 0.146. The maximum p-value of the correlation coefficient of each time lag at the 95% level of confidence was 7.91×10^{-10} . Since the D-W test statistic value is less than 2, the time series of daily calorific value reduction is positively autocorrelated. The autocorrelation function was also calculated for 100 days.

Using the statsmodels library for python [7] autocorrelation was analyzed for each day up to 100 days. The results are shown in Figure 9. The autocorrelation of the current day is always 1. The results located in the shaded area are significant ($\alpha = 0.05$). The autocorrelation coefficient falls to and remains less than 0.5 after a 1-day lag and there is a small positive or negative autocorrelation of the time series data after day 16.

The low VIF values shown in Table 2 indicate that the predictors are not correlated and consequently none of the coefficients are redundant and all the coefficients can be used in the polynomial regression model. Additionally, this result shows the independence of the predictors and their suitability for use in the dataset used to train and validate the ANN models.

The observation that the response data are autocorrelated implies that the data might have time dependent features. Consequently, we decided to further study time series features by using an LSTM model and by assessing the effect of data shuffling prior to training. Additionally to further evaluate time dependency, separate datasets were built using both shuffled and unshuffled data for training and testing. If time series features are important, then the LSTM model is expected to lead to more accurate predictions than the polynomial and the MLP models.

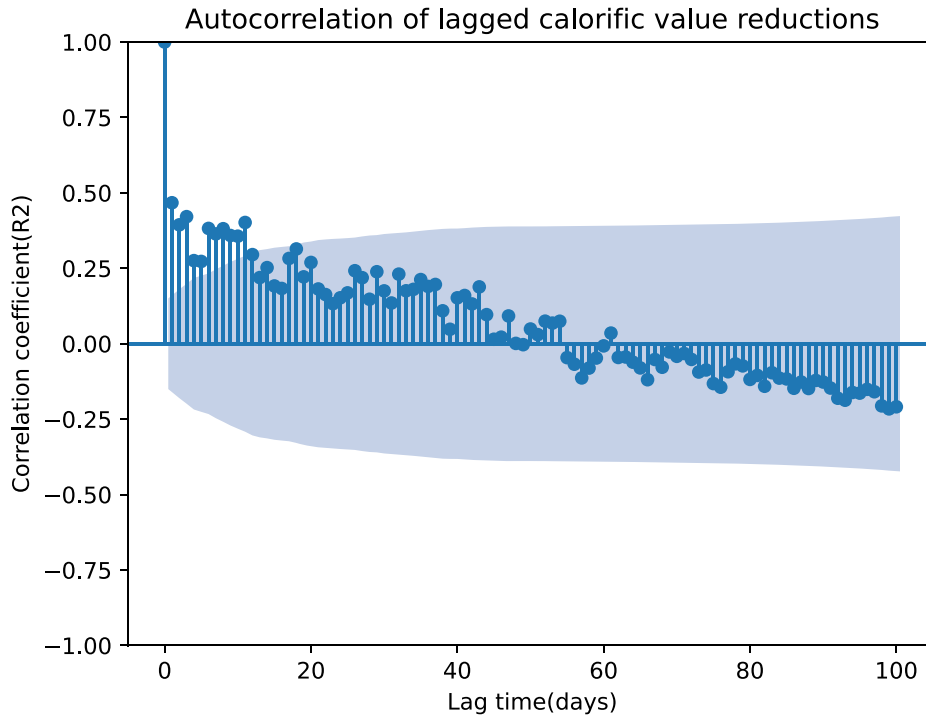


Figure 9: **Autocorrelation of lagged calorific value reduction** The lags are plotted on the horizontal and the correlations are plotted on the vertical axis.

Table 2: Variance Inflation Factors (VIF) for each predictor. *ESD*: equivalent spherical diameter; *MAT*: packing material; *HDR*: bed height/diameter ratio; *Q*: Influent flow rate.

| <i>VIF</i> | |
|------------|------|
| ESD | 2.31 |
| MAT | 2.05 |
| HDR | 2.80 |
| Q | 3.85 |

Polynomial model

The polynomial model was made using the sklearn linear_model package. The leave-one-out method was used to randomly remove a set of predictors and responses during each replication. The equation is shown below where Wr is the vector of model coefficients and Ir is the y-intercept.

$$\hat{y}_{poly} = X_{test_mm}[k,0]*Wr[0] + X_{test_mm}[k,1]*Wr[1] + X_{test_mm}[k,2]*Wr[2] + X_{test_mm}[k,3]*Wr[3] + Ir$$

MLP model

Each hidden layer unit input was propagated to every unit of the subsequent hidden layer dense units using the following propagation rule:

$$s_k^p = \sum_{i=1}^L w_{ik} y_i^p + b_k$$

where s_k^p is the vector propagated to unit k of the subsequent layer, L is the number of hidden layer units, w_{ik} is the value of the weight of the connection between unit i of the current layer and unit k of the subsequent layer, y_i^p is the output of the unit, and b_k is the value of the bias of this connection [1].

The error during training was calculated with the Mean Squared Error (MSE) loss function.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Where N, is the training batch size, y_i is the true value and \hat{y}_i is the predicted value. The architecture of the MLP ANN that was used for subsequent simulations is shown in Figure 10 below. Note: The final model had 4 hidden layers with 256 units per hidden layer.

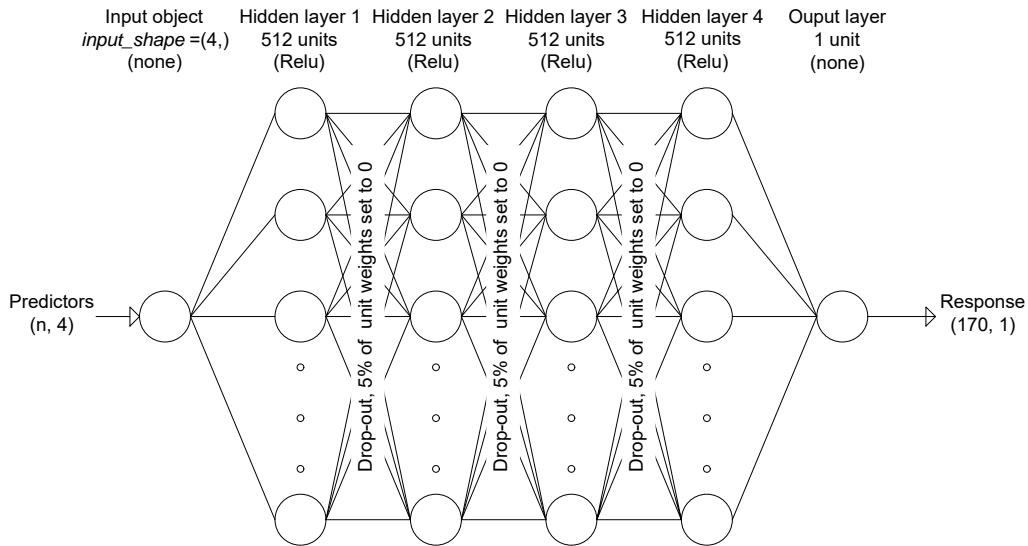


Figure 10: Architecture of the x-layer MLP used for simulation

Choosing the number of MLP hidden layers and units

To determine the number of hidden layers and hidden layer units, all possible MLP architectures having from 1 to 13 hidden layers and from 8 to 4096 units per hidden layer were built, trained and evaluated. The

prediction accuracy of the validation tests was evaluated in terms of the coefficient of determination (R^2) and the slope of the regression line. A value of 1 of both the coefficient of determination (R^2) and the slope indicates the best fit of the predicted to the true values. Consequently, a high value of the sum of R^2 and the slope indicates high accuracy. The results are shown in Figure 11 and in Figure 12. Model building, training and evaluation was repeated at least 10 times each with random initialisation of the model weights. We observed that the models having between 64 and 2048 units per layer had the highest accuracy and that using between 2 and 6 hidden layers yields similar results.

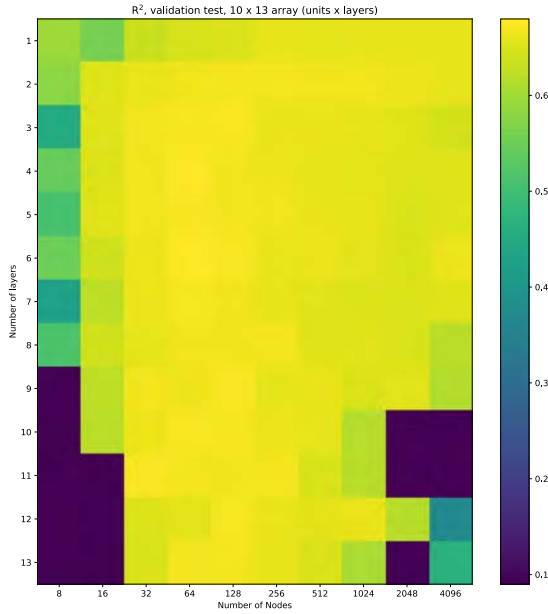


Figure 11: Validation test R-squared for different numbers of hidden layers and numbers of units per hidden layer

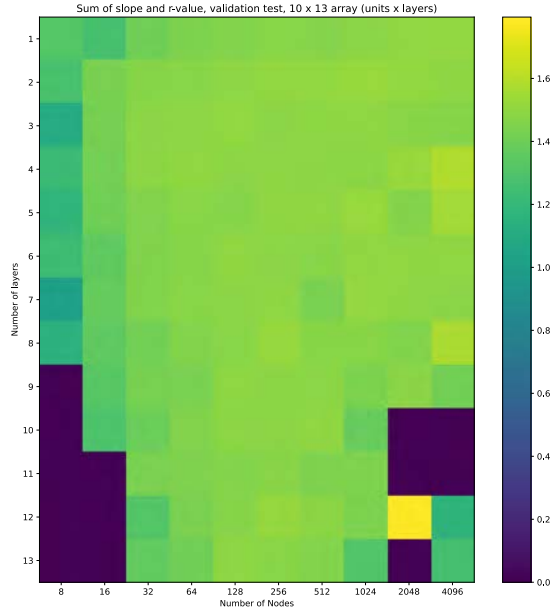


Figure 12: Validation test sum of R-squared and slope of the regression line for different numbers of hidden layers and numbers of units per hidden layer

LSTM model

In addition to the forward and backward propagation of errors that is a characteristic of all neural networks, recurrent neural networks (RNN) have feedback connections between neurons. Having feedback connections makes it possible to model persistent relations between data objects in a time series thereby giving RNNs the capacity for memory. The LSTM neural network is an improved class of RNN that has memory cells that feature input and output gates that protect the memory cell from perturbations due to, respectively, currently irrelevant inputs and memory cell contents [3]. Neural network architectures containing LSTM cells have demonstrated excellent ability for use in predictive models of time series data sets. The LSTM model was built using Python 3.7.6 and the Keras 2.3.0 API running on top of the TensorFlow 2.0 library for machine learning. The sequential model architecture was used. The model input was defined as a 3D tensor with shape [batch size, timestep, features] where *batch_size* refers to the number of days input to the model before updating the LSTM memory cell weights, *timestep* is the number of equal length sequences of data making up a single batch, and *features* refers to the 4 predictors (ESD, MAT, HDR, Q). One epoch was defined as a passage of all 1530 samples (days). The input tensor was fed directly to an LSTM layer comprised of 128 units. Following the Keras recommendations, the LSTM layer unit used the *tanh* activation function and the *sigmoid* recurrent activation function. The memory cell layer was updated between mini-batches of length *batch_size* as defined by the input shape definition. The *stateful* argument was set to *True* so that the layer weights were used as initial weights in the subsequent mini-batch. The *return_sequences* argument was set to *False* because the model contained a single hidden layer. The remaining hyperparameter arguments were set to default values. The output layer had a single *Dense* unit without activation.

Since LSTM layers are capable of using features that vary over long time periods, we tested input *batch_size* arguments between 1 and 170 days during training. The compiled regression models were trained using a mini-batch gradient descent back propagation learning algorithm. The error during training was

calculated with the Mean Squared Error (MSE) loss function. For the purpose of comparison, separate models were built using both shuffled and unshuffled data for training and testing.

I found that using unshuffled time-series data and a memory cell training `time_step` of 10 days yielded the best results.

The input and output gate activation functions adjust independently of the real rate and values of error information flow through the memory cell. Memory cell learning is the result of automatic scaling of the values by the input and output gates during model training. Having gates makes it possible to superimpose, onto the real-time error flow, selected errors from previous time steps that might be important to accurately modelling the time series data. An update layer combines the input and output gate vector values with the cell state. These functions make it possible to account for time series features of the data such as long time lags.

The LSTM cell is described in the figure 13 and the text below adapted from Olah ([5]).

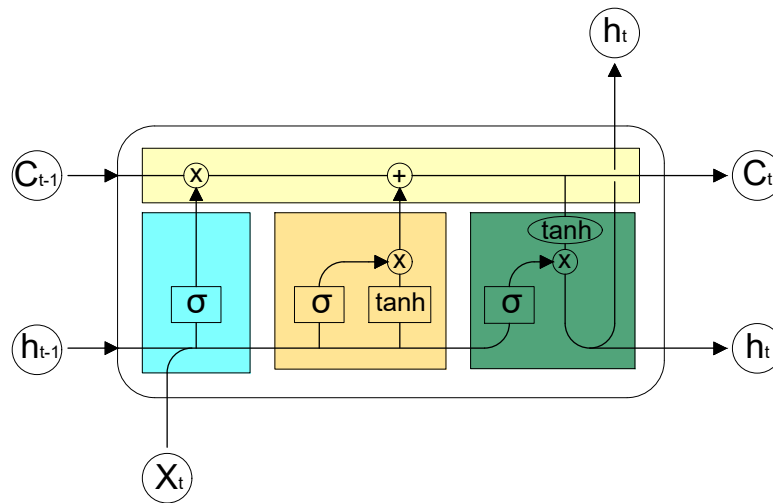


Figure 13: LSTM memory cell

The current data input and hidden state vectors are represented by x_t and h_t respectively. The current LSTM cell input vector length is a fraction of the sample length as specified in the model input shape definition. The cell state, also known as the constant error carousel (CEC), is represented by the yellow box. The automatic modifications to the cell state control the information contained in the output vectors. The cell state and hidden state vectors of the previous step are represented by C_{t-1} and h_{t-1} respectively.

The forget gate layer, represented by the blue shaded area, decides what information to keep from the previous cell state. This is done by multiplying each value of the cell state by the value returned by the *sigmoid* function of the updated weights and biases as follows:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)$$

The update layer, represented by the orange shaded area, combines a list of values to update (i_t) with the vector of new cell state values (\tilde{C}_t) determined using the *tanh* function of the updated weights and biases

as follows:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i)$$
$$\tilde{C}_t = \tanh(W_c * [h_{t-1}, x_t] + b_c)$$

The output layer, represented by the green shaded area, combines a list of values to update (o_t) with the \tanh function of the vector of the new cell state values (\tilde{C}_t) as follows:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * \tanh(C_t)$$

The vector of new cell state values (C_t) is determined as follows:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Thus, the memory cell updates after a defined number of timesteps and outputs a vector of scaled predictions.

3 Results

The reference calorific value reductions were compared to the predicted calorific value reductions. The predictors used in mock experiment number 3 are the same as those used in the reference experiment. Consequently, the calorific value reduction should be the same. As shown in figure 14 the linear relationship between the calorific value reductions in the reference data set and the mock data set demonstrates the capacity of the equation used to generate mock data to reproduce the response data acquired during the reference experiment.

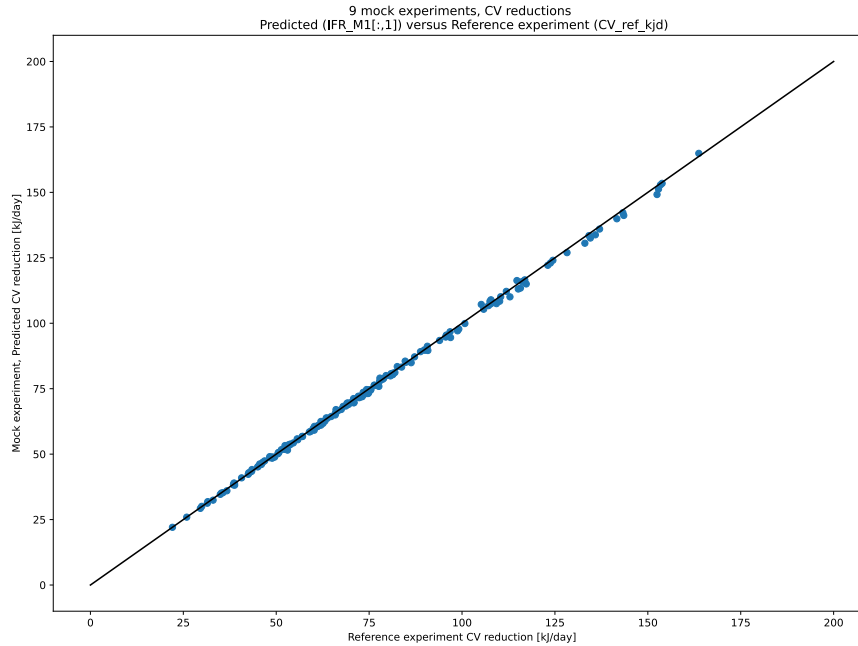


Figure 14: Calorific value reduction: reference data set versus mock experiment E3

As shown in figure 15 the rank of the experiments in terms of CV reduction varies with the flow rate. The predictors used in experiment E3 (ESD=36, MAT=PUF, HDR=3.6) yield the highest CV reduction when the flow rate is high. The predictors used in experiment E1 (ESD=4, MAT=PVC, HDR=0.5) yield the highest CV reduction when the flow rate is low. This implies that set of predictors can be selected according to the flow rate range.

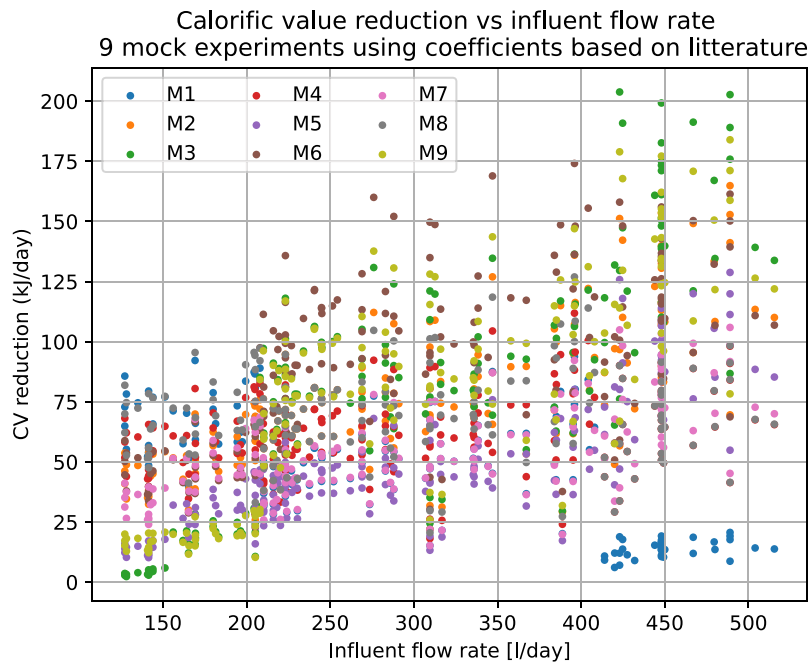


Figure 15: Calorific value reduction versus influent flow

Polynomial

As shown in figure 16 (unshuffled, true time series) the polynomial model does not accurately predict extreme values during the entire experiment.

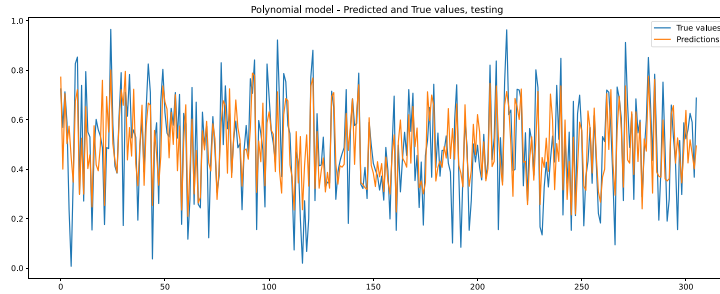


Figure 16

As shown in figure 17 (shuffled time series) the polynomial model does not accurately predict extreme values during the entire experiment.

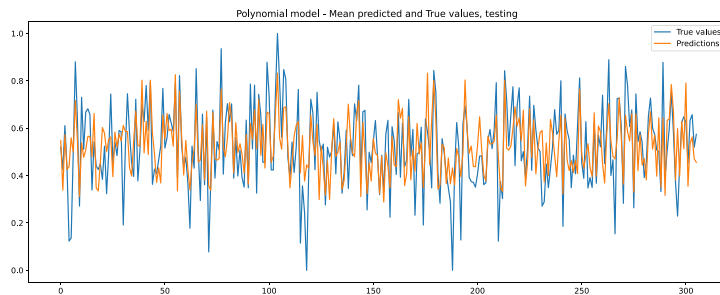


Figure 17

Comparison of the polynomial, MPL and LSTM models

Figure ?? shows plots of predicted versus true values obtained from the polynomial, MLP, and the LSTM models using not-shuffled, true time series data.

Figure ?? shows plots of predicted versus true values obtained from the polynomial, MLP, and the LSTM models using pre-shuffled data.

Figure 20 compares the results of analysis of the prediction accuracy of multiple runs of the polynomial model and of multiple runs of the MLP and LSTM models with random weight initialisation. The high slope and the narrow distribution of the slope and the RMSE demonstrate the accuracy of MLP model. The highway LSTM has a higher slope and RMSE and wider distribution of both slope and RMSE than the 2 other models. The polynomial model has a much lower slope and a higher RMSE than the MLP model.

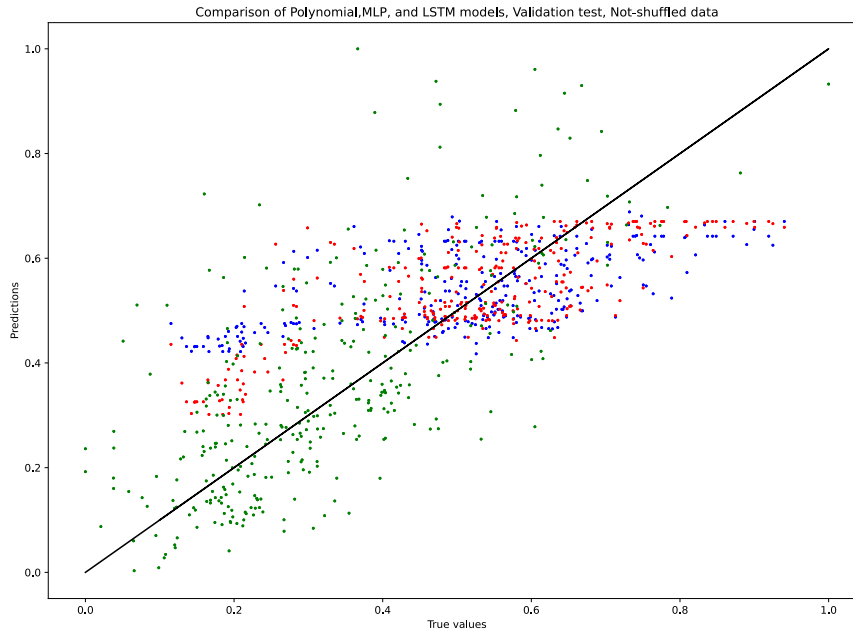


Figure 18: True versus predicted values (unshuffled, true times series data) Polynomial model: blue; MLP model: red; LSTM model: green

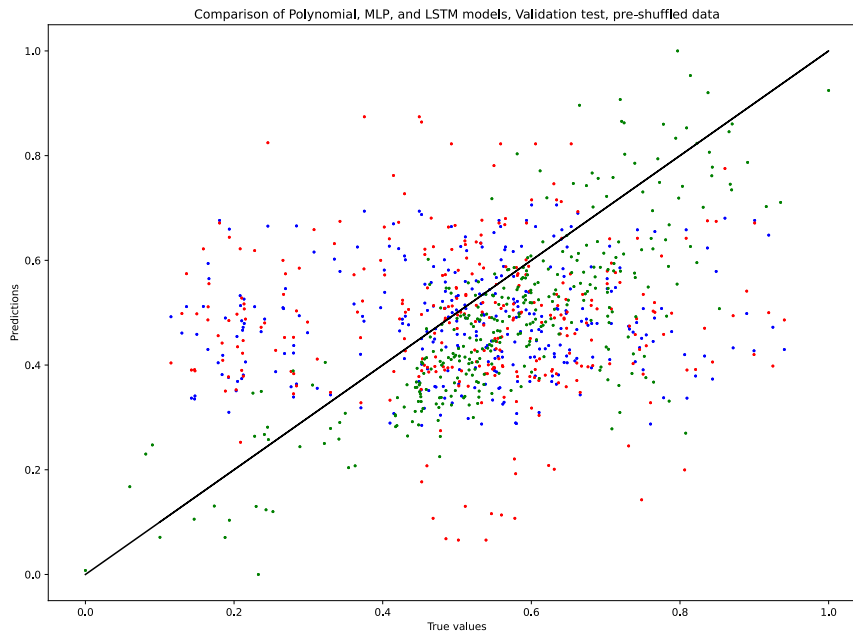


Figure 19: True versus predicted values (pre-shuffled data) Polynomial model:blue; MLP model: red; LSTM model: green

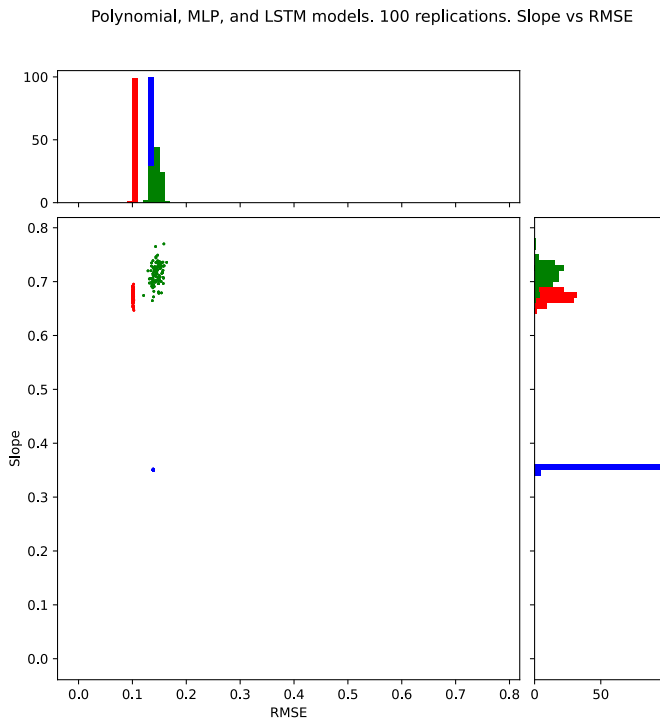


Figure 20: RMSE and Regression line slope of the polynomial model and multiple runs of the MLP and LSTM models . A slope equal to 1 corresponds to a perfect match between predicted and true values. The histograms show the corresponding frequency distributions. Polynomial model: blue; MLP model: red; LSTM model: green

Simulations

Figure 21 shows examples of the low (127 - 321 l/day) and high (322 - 516 l/day) range influent flow regimes used during simulation.

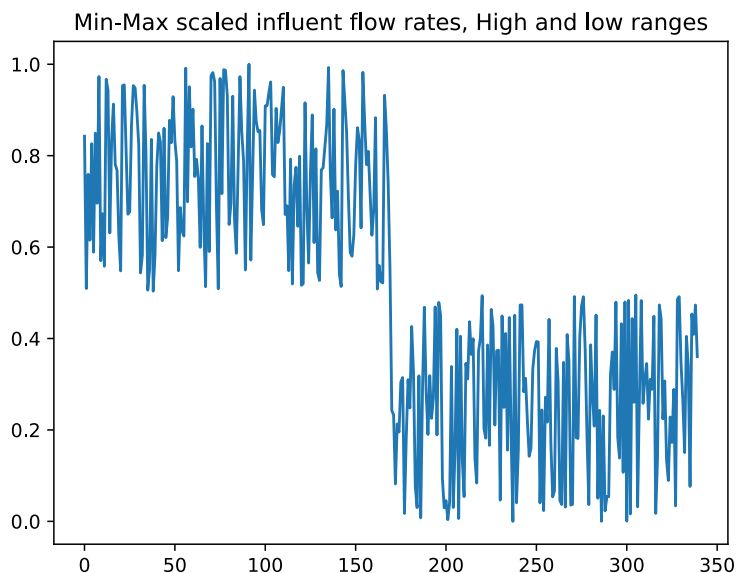


Figure 21: **Randomly generated influent flow rates.** Left: High range (322 - 516 l/day); Right: Low range (127 - 321 l/day).

References

- [1] P. Araujo et al. "Multilayer perceptron neural network for flow prediction". In: *J. Environ. Monit.* 13.1 (2011), pp. 35–41. DOI: 10.1039/C0EM00478B.
- [2] S Di Berardino, S Costa, and A Converti. "Semi-continuous anaerobic digestion of a food industry wastewater in an anaerobic filter". In: *Bioresource Technology* 71.3 (2000), pp. 261–266. DOI: [https://doi.org/10.1016/S0960-8524\(99\)00080-2](https://doi.org/10.1016/S0960-8524(99)00080-2).
- [3] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (1997), pp. 1735–1780.
- [4] D.J. Elliot Lynn C. Smith and A. James. "Mixing in upflow anaerobic filters and its influence on performance and scale-up". In: *Water Research* 30.12 (1996), pp. 3061–3073. DOI: 10.1016/S0043-1354(96)00169-8.
- [5] C. Olah. *Understanding LSTMs*. 2015. URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs>.
- [6] Jean Philippe Steyer Renaud Escudié Thierry Conte and Jean Philippe Delgenès. "Hydrodynamic and biokinetic models of an anaerobic fixed-bed reactor". In: *Process biochemistry* 40.7 (40 2005), pp. 2311–2323. DOI: doi:10.1016/j.procbio.2004.09.004.

- [7] Skipper Seabold and Josef Perktold. “statsmodels: Econometric and statistical modeling with python”. In: *9th Python in Science Conference*. 2010.
- [8] Dhandapani Thirumurthi. “Effects of mixing velocity on anaerobic fixed film reactors”. In: *Water Research* 22.4 (1988), pp. 517–523. DOI: 10.1016/0043-1354(88)90049-8.

Exploratory study of municipal wastewater primary sedimentation basin effluent treatment using an upflow anaerobic filter

Author: Mark McCormick, Environmental Engineer, M.Sc., Tel: +41 78 604 52 42,
mark_mccormick@hotmail.com

Revision date: 2021.10.29

Abstract

This report describes the results of exploratory tests conducted using a 95-liter upflow anaerobic filter (UAF) bioreactor test unit fed raw primary sedimentation basin effluent at a municipal wastewater treatment plant. The main objectives of the exploratory study were to obtain substrate conversion and hydraulic data for use in making full-scale projections, evaluate potential biofilm support materials, and to assess the suitability of the upflow anaerobic filter unit process as part of a new concept for a complete municipal wastewater treatment plant. The tests were conducted using available equipment with no effort made to optimize the system. The main findings of the tests are: COD loading rate: 2,1 kg COD/m³/day, hydraulic retention time: 3 hours, average influent energy conversion: 71% (Influent HHV -effluent HHV), biogas CH₄ content up to 65%, stable operation at 16°C, evaluation of different biofilm support materials (see text), non-biodegradable ash and phenolic compounds account for at least 63% of the total solids contained in primary sedimentation basin effluent and they are the most important contributors to turbidity.

Introduction

This work was conducted as part of an effort to conceive a new wastewater treatment and resource recovery system to replace the conventional wastewater treatment plant designs. The aim is to achieve optimal conception of the system and integration of all unit processes to achieve minimal construction and operating cost, and maximal efficiency, profit, and environmental quality.

Some of the major problems inherent to conventional municipal wastewater treatment are:

- Production of excess activated sludge
- Recalcitrance to anaerobic digestion of activated sludge compared to primary sludge
- More than half of the available chemical potential energy is lost as CO₂ released from the activated sludge and the anaerobic digestion processes
- Concentration of recalcitrant substances due to internal recycle
- All the produced electricity is consumed internally (none is profitably exported)
- Not all heat is used productively
- Requirement to add consumable chemicals to remove phosphorous and improve sludge sedimentation

In industrialized countries, the cost of wastewater treatment is affordable in terms of economic cost and energy consumption. Nevertheless, if conventional activated sludge processes were replaced by the proposed anaerobic UAF process, then wastewater treatment would be an energy producing activity and electricity could be exported from wastewater treatment plants and carbon emissions from fossil fuel consumption could be avoided. In the case of lower income countries, high operating costs, especially electricity purchases, might explain the current situation where more than 2,5 billion people are not served by wastewater treatment plants and billions more are poorly served by inadequate wastewater treatment.

Conventional wastewater treatment relies on aeration and on aerobic microorganisms to oxidize the organic fraction of municipal wastewater to CO₂ and to produce sludge that can be removed from water by decantation. This strategy and the first activated sludge processes were first implemented over 100 years ago. Nitrogen and phosphorous are removed by additional aerobic processes or by chemical precipitation. Although effective, conventional wastewater treatment requires approximately 7.5 W/capita of electrical power input, mostly for aeration. The total electricity requirement for wastewater treatment in Switzerland is approximately 500 GWh per year. Up to 80% of this requirement can be met by on-site Combined Heat and Power (CHP) production using biogas fuel. Nevertheless, conventional wastewater treatment is very energy inefficient. In a typical wastewater treatment plant over 50% of the calorific value of the incoming wastewater solids is lost as CO₂ produced by bacterial respiration. Moreover, the excess thermal energy produced by CHP is often not used and no electricity is exported to the grid. Energy flows from a typical activated sludge based wastewater treatment plant are described in the diagram below.

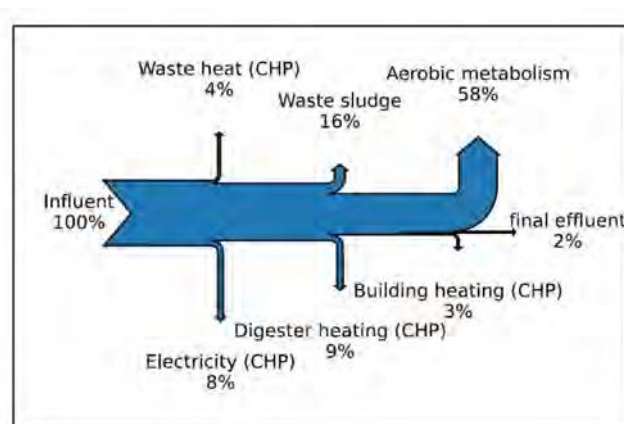


Figure 1 Energy balance of a conventional Activated Sludge WWTTP

The upflow anaerobic filter (UAF) is an anaerobic digester that contains a packed bed of solid supports on which a biofilm develops. Influent enters below the filter and exits above the filter.

This study was conceived to explore the use of an UAF for the treatment of municipal wastewater primary effluent in real industrial conditions. The test equipment was recovered from previous projects and adapted to build the anaerobic filter with a very small budget for materials. This approach made it possible to acquire preliminary results that can be used to assess the feasibility and to orient the work on the preparation of new project proposals to make detailed designs of anaerobic filters and to develop new complete wastewater treatment plant system concepts.

The main research topics and questions addressed by this study were:

- What are the chemical conversion rates (biogas and VFA production, COD changes, etc)?
- What is the maximum loading rate?
- What is the minimum hydraulic retention time?
- Biofilm characteristics
- What materials should be used for biofilm supports?
- Material and energy balance
- Temperature effects
- Residence time distribution
- Effects on downstream denitrification
- Scale-up potential

The quality of the results obtained during this exploratory study is not adequate for submission to a scientific journal. Nevertheless, they are reported here because they might contribute to efforts to improve urban wastewater treatment systems and technology.

Material and methods

A simple test unit was constructed to collect data continuously for a long duration under a wide variety of industrial operating conditions. Unfiltered, unheated UAF influent was taken directly from the wwtp primary sedimentation basin effluent and fed continuously to the test unit. The flow rate was regulated by a peristaltic pump. Temperature, biogas flow rate, CH₄, pH and ORP were measured at appropriate intervals to permit continuous recording of operational data. A post digester decanter served as a gas/liquid/solid separator.

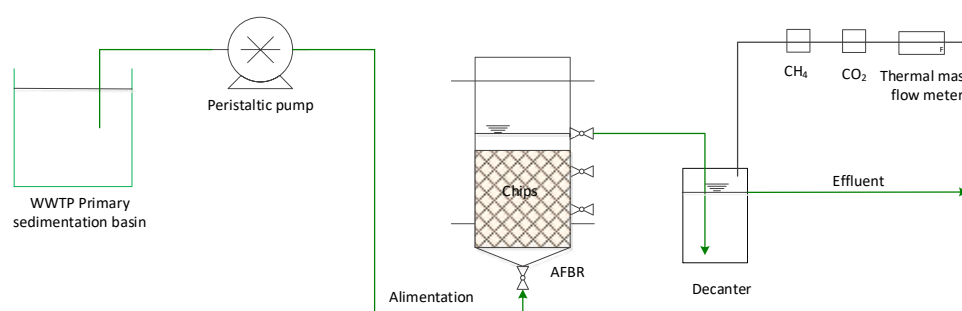


Figure 2 UAF test unit



| Anaerobic filter | litres |
|----------------------------|--------|
| Total volume | 79 |
| Gas head space | 23 |
| Liquid volume (water) | 37 |
| Support volume (torrefied) | 19 |
| Total working volume | 56 |
| Decanter (post bioreactor) | 16 |

Figure 3 UAF test unit (from left to right: inlet, test unit, compartment volumes)

From May 4 to December 15th 2017 (215 days) this test unit was operated almost continuously to treat primary sedimentation basin effluent from the municipal wastewater treatment plant in Yverdon-les-Bains, Switzerland.

After conducting tests for more than 7 months at the wastewater treatment plant, the test unit was moved to the laboratory for residence time distribution studies using tap water with salt added to increase conductivity (described below in the results section).

Anaerobic Filter Bed description

The filter bed material was produced by heating wood chips at 250°C in an anoxic atmosphere (torrefaction). The average dimensions of the torrefied chips were 28 x 8 x 4 mm. 30 liters of biofilm coated torrefied chips from a previous experiment were mixed with 12 liters of clean torrefied chips and loaded into the bioreactor. In order to investigate the use of other biofilm supports, different plastic test materials were placed in perforated containers that were buried in the torrefied chip anaerobic filter bed. The torrefied wood chips and the plastic samples were removed periodically, and the activity of the associated biofilm was evaluated.

Results

The measured parameters are shown in the table below and in figures in the text. Additionally, selected results are presented and commented below.

Summary of major process parameters

Samples were collected, and off-line measurements of the usual anaerobic digestion process parameters were made at 1 to 3-day intervals. Some parameters were measured weekly or even less frequently. Additional results are presented graphically at the end of this report and more are available on request.

Table 1 Summary of influent off-line measurements

| Parameter | Unit | Number | Mean | standard deviation | RSD |
|-------------------------------------|-----------|--------|--------|--------------------|------|
| Inlet flow rate | l/jour | 212 | 268 | 114 | 0.42 |
| Inlet water temperature | °C | 118 | 19.10 | 7.51 | 0.39 |
| pH | | 109 | 7.45 | 0.22 | 0.03 |
| Total solids | % | 24 | 0.06 | 0.02 | 0.24 |
| Total volatile solids | % of TS | 23 | 37.31 | 8.36 | 0.22 |
| Ash | % of TS | 23 | 62.80 | 8.51 | 0.14 |
| HHV | MJ/kg dry | 10 | 5.96 | 0.87 | 0.15 |
| LHV | MJ/kg dry | 9 | 5.48 | 0.88 | 0.16 |
| C | % TS | 10 | 19.08 | 1.30 | 0.07 |
| H | % TS | 10 | 2.37 | 0.24 | 0.10 |
| N | % TS | 10 | 1.58 | 0.19 | 0.12 |
| S | %TS | 1 | 0.49 | | |
| O | % TS | 10 | 14.2 | 1.41 | 0.02 |
| tATP (total Adenosine triphosphate) | ng/l | 5 | 68.54 | 45.97 | 0.67 |
| VFA, expressed as acetate | mg/l | 26 | 94.71 | 29.28 | 0.31 |
| COD, dissolved | mg/l | 13 | 255.74 | 90.92 | 0.36 |
| COD, suspended | mg/l | 13 | 429.46 | 144.95 | 0.34 |
| COD, suspended (filtered 0,45 µm) | mg/l | 4 | 264.50 | 38.53 | 0.15 |
| COD, dissolved (filtered 0,45 µm) | mg/l | 4 | 276.50 | 36.59 | 0.13 |
| TOC, dissolved | mg/l | 10 | 80.70 | 29.69 | 0.37 |
| TOC, suspended | mg/l | 8 | 123.88 | 16.30 | 0.13 |
| TOC, suspended (filtered 0,45 µm) | mg/l | 2 | 59.90 | 19.66 | 0.33 |
| TKN (LatoN) dissolved | mg/l | 12 | 67.84 | 25.82 | 0.38 |
| TKN (LatoN) suspended | mg/l | 4 | 62.63 | 23.09 | 0.37 |
| NH4-N | mg/l | 26 | 53.12 | 21.38 | 0.40 |
| NO3-N | mg/l | 9 | 0.63 | 0.21 | 0.34 |
| PO4 | mg/l | 16 | 5.79 | 3.43 | 0.59 |
| SO4 dissolved | mg/l | 3 | 68.60 | 7.10 | 0.10 |
| SO4 suspended | mg/l | 1 | 93.40 | | |
| Absorbance, 230 nm | A.U. | 12 | 0.01 | 0.02 | 2.10 |
| Absorbance, centrifuged, 230 nm | A.U. | 9 | 0.01 | 0.02 | 2.24 |
| Absorbance, 260 nm | A.U. | 28 | 0.88 | 0.60 | 0.68 |
| Absorbance, centrifuged, 260 nm | A.U. | 9 | 0.26 | 0.06 | 0.24 |
| Absorbance, 280 nm | A.U. | 12 | 0.63 | 0.31 | 0.49 |
| Absorbance, centrifuged, 280 nm | A.U. | 9 | 0.34 | 0.08 | 0.23 |
| Absorbance, 830 nm | A.U. | 28 | 0.25 | 0.16 | 0.64 |

| | | | | | |
|---------------------------------|-------|-----|------|------|------|
| Absorbance, centrifuged, 830 nm | A.U. | 9 | 0.10 | 0.04 | 0.44 |
| Absorbance, 860 nm | A.U. | 28 | 0.24 | 0.16 | 0.65 |
| Absorbance, centrifuged, 860 nm | A.U. | 9 | 0.08 | 0.03 | 0.42 |
| Absorbance, 890 nm | A.U. | 27 | 0.23 | 0.15 | 0.67 |
| Absorbance, centrifuged, 890 nm | A.U. | 9 | 0.07 | 0.03 | 0.43 |
| HRT | hours | 209 | 4.66 | 5.69 | 1.22 |

Table 2 Summary of effluent off-line measurements

| Parameter | Unit | Number | Mean | standard deviation | RSD |
|--------------------------------------|-----------|--------|--------|--------------------|------|
| pH (digester effluent) | | 29 | 7.14 | 0.16 | 0.02 |
| pH (post digester decanter effluent) | | 104 | 7.16 | 0.18 | 0.02 |
| Temperature (decanter effluent) | °C | 110 | 22.75 | 6.83 | 0.30 |
| Total solids | g/l | 24 | 0.05 | 0.01 | 0.20 |
| Total volatile solids | % TS | 23 | 29.05 | 6.71 | 0.23 |
| Ash | % TS | 23 | 70.95 | 6.71 | 0.09 |
| HHV | MJ/kg dry | 10 | 2.31 | 1.06 | 0.46 |
| LHV | MJ/kg dry | 9 | 2.10 | 1.04 | 0.50 |
| C | % TS | 10 | 13.49 | 1.73 | 0.13 |
| H | % TS | 10 | 1.13 | 0.45 | 0.40 |
| N | % TS | 10 | 0.97 | 0.27 | 0.28 |
| O | % TS | 10 | 14.0 | 2.40 | 0.03 |
| tATP (total Adenosine triphosphate) | ng/l | 4 | 70.97 | 42.19 | 0.59 |
| VFA, expressed as acetate | mg/l | 25 | 55.37 | 24.68 | 0.45 |
| COD, dissolved | mg/l | 16 | 124.23 | 50.66 | 0.41 |
| COD, suspended | mg/l | 10 | 221.82 | 58.69 | 0.26 |
| COD, suspended (filtered 0,45 µm) | | 4 | 143.00 | 25.26 | 0.18 |
| COD, dissolved (filtered 0,45 µm) | mg/l | 4 | 137.25 | 12.26 | 0.09 |
| TOC, dissolved | mg/l | 11 | 42.95 | 11.90 | 0.28 |
| TOC, suspended | mg/l | 9 | 66.60 | 16.74 | 0.25 |
| TOC, suspended (filtered 0,45 µm) | | 2 | 28.45 | 9.26 | 0.33 |
| TKN (Laton) dissolved | mg N/l | 10 | 45.81 | 13.58 | 0.30 |
| TKN (Laton) suspended | mg N/l | 4 | 56.88 | 29.11 | 0.51 |
| NH4-N dissolved | mg N/l | 26 | 39.54 | 11.56 | 0.29 |
| NO3-N | mg N/l | 8 | 0.42 | 0.09 | 0.23 |
| PO4 | mg/l | 17 | 5.42 | 1.81 | 0.33 |
| SO4 dissolved | mg/l | 4 | 32.35 | 12.16 | 0.38 |
| SO4 suspended | mg/l | 1 | 52.40 | | |
| Absorbance, 230 nm | A.U. | 12 | 0.01 | 0.02 | 2.12 |
| Absorbance, centrifuged, 230 nm | A.U. | 9 | 0.01 | 0.02 | 2.24 |
| Absorbance, 260 nm | A.U. | 28 | 0.57 | 0.31 | 0.54 |
| Absorbance, centrifuged, 260 nm | A.U. | 9 | 0.21 | 0.06 | 0.28 |
| Absorbance, 280 nm | A.U. | 12 | 0.40 | 0.15 | 0.38 |
| Absorbance, centrifuged, 280 nm | A.U. | 9 | 0.22 | 0.07 | 0.30 |
| Absorbance, 830 nm | A.U. | 28 | 0.14 | 0.06 | 0.43 |
| Absorbance, centrifuged, 830 nm | A.U. | 9 | 0.07 | 0.03 | 0.41 |
| Absorbance, 860 nm | A.U. | 28 | 0.13 | 0.06 | 0.43 |
| Absorbance, centrifuged, 860 nm | A.U. | 9 | 0.06 | 0.03 | 0.41 |
| Absorbance, 890 nm | A.U. | 27 | 0.13 | 0.06 | 0.44 |
| Absorbance, centrifuged, 890 nm | A.U. | 9 | 0.06 | 0.02 | 0.43 |

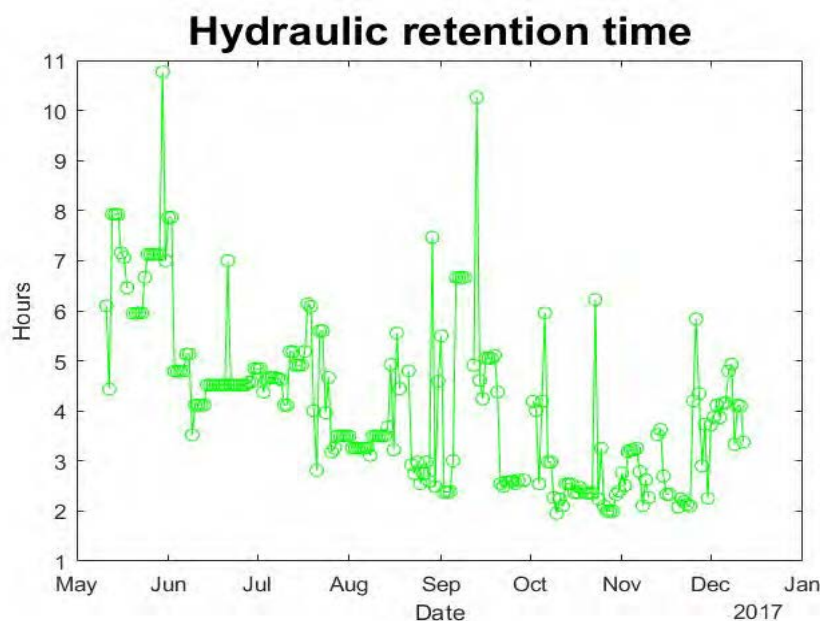


Figure 4 Hydraulic retention time

Solids removal

The anaerobic filter removed both dissolved and suspended COD. Note that primary sedimentation basin effluent contained more than 60% ash (non-volatile fraction of total solids).

Table 1 Solids removal

| Anaerobic filter | | Total solids | Volatile solids | COD, dissolved | COD, suspended | COD, total |
|------------------|------|--------------|-----------------|----------------|----------------|------------|
| Influent | mg/l | 600 | 222 | 256 | 429 | 685 |
| Effluent | mg/l | 500 | 145 | 124 | 223 | 347 |
| Reduction | % | 17 | 35 | 52 | 48 | 49 |

Considering that primary sedimentation contains a large fraction of inorganic materials, biological solids removal should be assessed by comparing UAF influent and effluent suspended TOC and suspended TKN. These organic fractions comprise large molecules such as proteins, fats and cellulose. The anaerobic filter removed 46% of the TOC in suspension and 9% of the TKN in suspension.

Tableau 1 Suspended COT and TKN removal

| Substance | | Dissolved TOC | Suspended TOC | Total TOC | NH4-N | NO3-N | NH4-N + NO3-N | Dissolved TKN | Suspended TKN | Total TKN |
|---------------------------------------|----------|---------------|---------------|-----------|-------|-------|---------------|---------------|---------------|-----------|
| Balance (UAF effluent – UAF influent) | mmoles/l | -3,1 | -4,8 | -7,9 | -1,0 | 0,0 | -1,0 | -1,6 | -0,4 | -2,0 |
| Balance (UAF effluent – UAF influent) | % | -46,8 | -46,2 | -46,5 | -25,6 | -33,3 | -25,7 | -32,5 | -9,2 | -21,3 |

In addition to comparing influent and effluent total solids determined gravimetrically, it is useful to simply observe the dried solids. Influent solids were a dark brown coloured viscous paste. Effluent solids were a light brown coloured sand like powder. As shown in the photo below, solids are removed

by the anaerobic filter. The solids remaining in the effluent are 71 % non-volatile. The ash fraction cannot be removed biologically. The non-volatile fraction is apparently recalcitrant to biological degradation.



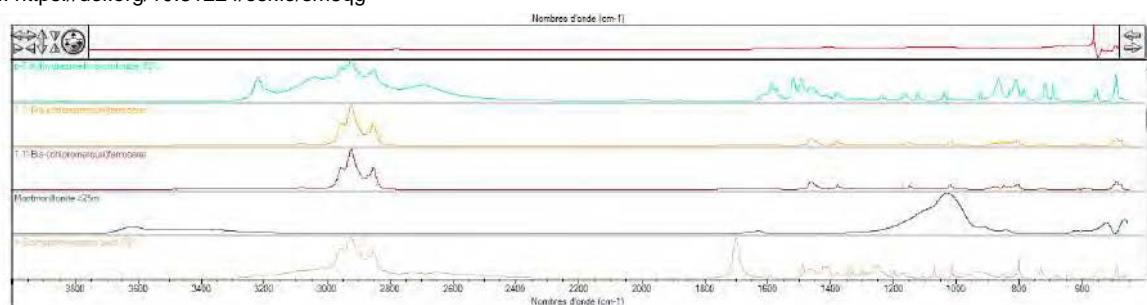
Figure 5 Dried influent (left), dried effluent (right)

FT-IR analysis of total solids

Influent and effluent liquid samples were dried at 105°C. The dried solids were ground to obtain a fine powder. FT-IR spectra of the dried solids were obtained and compared to spectral libraries to identify substances present in the dried solids. The method is qualitative and not quantitative. The identified substances can be classified into 3 broad categories:

- Mineral clays
- Benzene and benzene derivatives
- Phosphonium derivatives

All three categories contain chlorinated derivatives. This exploratory study did not reveal a significant difference between the influent and the effluent. An example FT-IR spectrum is given in the figure below.



| | Correspondance | Titre | Dossier |
|---|----------------|-------------------------------------|--|
| 1 | 34.58 | Dichloroisopropylphosphine, 97% | HR Aldrich FT-IR Collection Edition II |
| 2 | 32.33 | Benzene-d6, 100.0 atom % D | HR Aldrich FT-IR Collection Edition II |
| 3 | 32.33 | Benzene-d6, 100.0 atom % D | HR Aldrich Organometallic, Inorganic, Silanes, Boranes, and Deuterated Compounds |
| 4 | 31.42 | p-Tolylhydrazine hydrochloride, 98% | HR Aldrich FT-IR Collection Edition II |
| 5 | 30.14 | 1,1'-Bis-(chloromercuri)ferrocene | HR Aldrich Organometallic, Inorganic, Silanes, Boranes, and Deuterated Compounds |

Figure 6 Substances identified in dried solids by FT-IR

Biofilm assessment

The presence of a biofilm and biofilm activity were assessed using three methods

- Aldehyde dehydrogenase activity
- ATP concentration
- Scanning electron microscopy

Aldehyde dehydrogenase activity

The oxidation of organic compounds is a dehydrogenation process mediated by specific dehydrogenase enzymes. Therefore, The Dehydrogenase Enzyme Activity (DHA) reflects microbial activity. The DHA assay is based on the reaction of redox-sensitive tetrazolium indicators. Triphenyl tetrazolium chloride (TTC) compounds are colourless and, after reduction, they are transformed into water insoluble, red coloured mono-formazans (TPF), which can be quantified by spectrophotometry. The DHA activity assay was adapted to the conditions of the AFBR project by a student intern and used to assess biofilm support materials including the torrefied chips, plastic foam and hard plastic supports. The samples were placed in the AFBR for between 15 and 98 days and then removed for assessment of biofilm activity. Biofilm specific activity can be expressed in terms of TPF/m² support surface. The highest biofilm specific activity was measured on the polyurethane foam supports. The activity measured using the ADH assay was 5 times greater on the polyurethane foam supports than the activity measured on the next best performing supports (some types of hard plastics and torrefied wood chips). The activity of biofilms that developed on torrefied wood was the same or higher than almost all the hard plastic supports.

ATP concentration and UV measurements

The concentration of ATP is positively correlated with the number of bacteria. The number of bacteria per volume of biofilm support can be evaluated by determination of ATP. Biofilm coated chips were

removed from the AFBR and ground to dislodge the bacteria. The total ATP concentration [tATP] of the suspension was measured and then divided by the measured volume of chips. The anaerobic filter biofilm concentration factor is defined as follows:

Biofilm concentration factor = tATP on supports/tATP in the bulk water

| Sample number | tATP [ng/ml] |
|------------------------------|--------------|
| 1 | 81.95 |
| 2 | 111.1 |
| 3 | 105.2 |
| Mean | 99.4 |
| | |
| AFBR Influent | 30.2 |
| AFBR Effluent | 31.4 |
| | |
| Biofilm concentration factor | 3.2 |

Table 2 total ATP determinations

Biofilm concentration factor = 99.4 [ng/ml]/ 30.8[ng/ml] = 3.2

The concentration of ATP measured on the biofilm support was 3.2 times greater than the concentration measured in both the influent and the effluent. The increase in the bacterial concentration between the influent and effluent streams was 4% (one measurement). This result demonstrates that a biofilm did develop on the torrefied wood chips and that it does not release bacteria to the effluent at a rate that is greater than the rate of bacteria introduction to the AFBR via the influent.

UV measurement

UV measurements at 280 nm give an indication of the protein content of the water and measurements at 260 nm indicate the DNA content. DNA content is correlated with the presence of microorganisms. Measurements of the influent and the effluent showed that both protein and DNA content decrease in the AFBR supporting the conclusion that the biofilm is attached to the filter bed and that the AFBR removes protein and DNA.

Scanning electron microscopy

The scanning electron micrographs made during a previous study conducted in the same AFBR also using torrefied wood chip biofilm supports but fed dairy wastewater and not municipal primary effluent are shown below. The biofilm is present after 4 months of continuous operation. However, the surface is not entirely covered. The surface coverage and consequently the total biological activity is expected to increase with longer incubation times and increased total biofilm surface coverage.

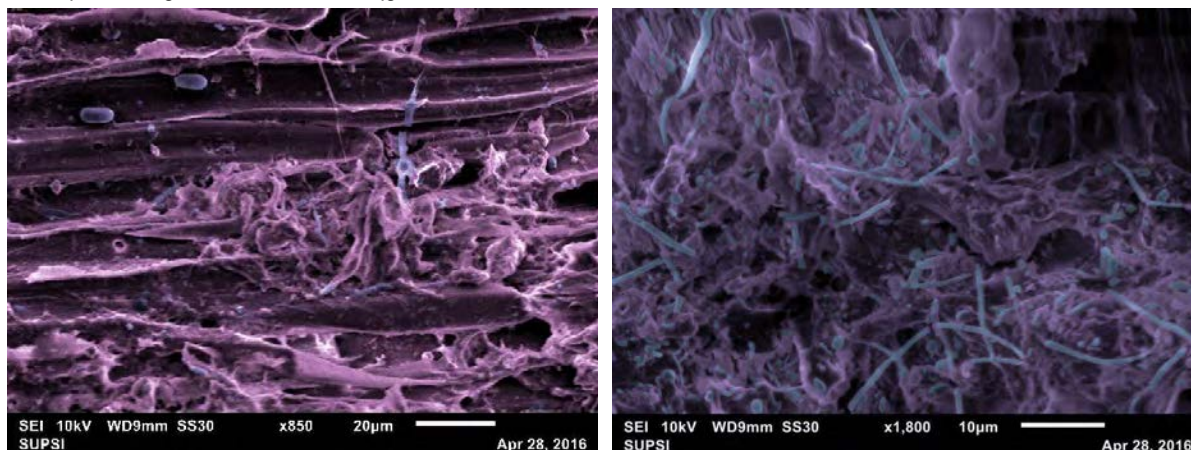
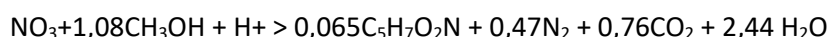


Figure 7 Scanning electron micrographs of biofilms on torrefied wood chips

Impact on carbon supply for denitrification

The loss of TOC in the UAF might result in insufficient dissolved TOC for the downstream denitrification process. This risk was evaluated by comparing the measured concentrations to the theoretical TOC requirement for denitrification.

The theoretical stoichiometric carbon requirement for denitrification is obtained from the equation below¹ and the measured TKN + NO₃. Prior conversion of NH₄-N to NO₃-N is assumed.



If the difference between the dissolved TOC and the theoretical carbon requirement and is less than zero, then carbon will have to be added to the denitrification process. The difference and the ratio of C/N in the primary sedimentation basin effluent (UAF influent) were evaluated and are presented in the table below.

Table 3 UAF influent TOC and N ratio

| Anaerobic filter (UAF) Stream characteristics | | Dissolved TOC | Suspended TOC | Total TOC | NH ₄ -N | NO ₃ -N | NH ₄ -N + NO ₃ -N | Dissolved TKN | Suspended TKN | Total TKN |
|--|-------------|---------------|---------------|-----------|--------------------|--------------------|---|---------------|---------------|-----------|
| | | Influent | mg/l | 80, | 123, | 204, | 53, | 0, | 53, | 67, |
| Influent | mmoles/l | 6,7 | 10,3 | 17,0 | 3,8 | 0, | 3,8 | 4,8 | 4,5 | 9,3 |
| TOC requirement for conversion | mmoles/l | | | | 4,1 | 0, | 4,1 | 5,2 | 4,8 | 10,1 |
| Balance (Influent dissolved TOC – Required TOC) | mmoles/l | 2,6 | | | | | | | | |
| Ratio C/N dissolved | moleC/moles | 1,4 | | | | | | | | |
| Ratio C/N total | moleC/moles | 1,8 | | | | | | | | |

Based on the averages of the entire study duration, the difference between the concentration of dissolved TOC in the influent and the TOC requirement for conversion of nitrate obtained from influent

¹ Metcalf & Eddy, Wastewater Engineering, Third edition, 1991

ammonium is 2,6 mmols/l. In the studied primary effluent stream, the balance is positive and the ratio is favorable for downstream denitrification.

However, as shown below, the UAF increases the risk of a lack of carbon for downstream denitrification. The difference between the concentration of dissolved TOC in the influent and the TOC requirement for conversion of nitrate obtained from influent ammonium is only 0.5 mmols/l. The reduced C/N ratios of the UAF effluent streams confirm this risk. This result is an important argument against implementation of UAF treatment of primary effluent in a conventional wastewater treatment plant with denitrification. As a pre-treatment implemented at some sites in a regional network, the impact of the carbon loss is reduced because some streams will not be treated by UAF.

Table 4 Impact on denitrification

| Downstream denitrification process | | COT dissoute | COT en suspension | COT totale | NH4-N | NO3-N | NH4-N + NO3-N | NKT dissoute | NKT en suspension | NKT totale |
|------------------------------------|---------------|--------------|-------------------|------------|-------|-------|---------------|--------------|-------------------|------------|
| Influent | mg/l | 43,0 | 66,6 | 109,6 | 39,5 | 0,4 | 40,0 | 45,8 | 56,9 | 102,7 |
| Influent | mmoles/l | 3,6 | 5,6 | 9,1 | 2,8 | 0,0 | 2,9 | 3,3 | 4,1 | 7,3 |
| Carbon requirement | mmoles/l | | | | 3,1 | 0,0 | 3,1 | 3,5 | 4,4 | 7,9 |
| Balance (TOCdis – C required) | mmoles/l | 0,5 | | | | | | | | |
| Ratio C/N dissolved | moleC/moles N | 1,1 | | | | | | | | |
| Ratio C/N total | moleC/moles N | 1,2 | | | | | | | | |

Digester temperature

Type K thermal couples were used to measure the temperatures of the:

- Inlet stream
- Centre of the anaerobic filter
- Outlet stream
- Ambient air

The AFBR digester is equipped with a concentric water jacket that was used to attempt to maintain a constant temperature inside the AFBR. However, it was not possible to cool the digester during hot summer days and it was not possible to adequately heat the digester when the flow rate was high and the water temperature was low.

The difference in temperature between the thermocouple placed at the center of the digester and a hand-held temperature probe used to control the temperature on 3 different days was between -0.8 and + 2.4 °C. Consequently, the reported temperature of the anaerobic filter bed (digester) is believed to be within 3° of the true value. The temperature set points during the experiment are shown in Table 5 Digester temperature set points:

| Dates | Number of days | Temperature set point [°C] | Average digester temperature [°C] |
|--|----------------|----------------------------|-----------------------------------|
| June 28 th to August 19 th | 52 | Not heated | 27 |
| August 20 th to September 19 th | 31 | 25°C | 23 |
| September 20 th to November 8 th | 50 | 20°C | 20 |
| November 9 th to December 15 th | 38 | Not heated | 13 |

Table 5 Digester temperature set points

One of the main results of this study was the demonstration of stable operation in different temperature intervals. Nevertheless, biogas production decreased towards the end of the study after the influent temperature decreased to less than 16°C. Between November 2017 and March 2018, the measured wwtp primary effluent temperature was most often between 14 and 20 °C. The temperature was below 15°C for only short durations and almost never was below 10°C. These results and observations suggest that the AFBR concept can be usefully applied to treat un-heated wastewater in temperate climates.

Biogas production

Both biogas and liquid effluent exited the bioreactor via the same tube that was connected to a closed decanter that served as a gas/liquid/solids separator. Biogas exited the decanter head space through a 4 mm tube. A very small headspace pressure of ~2 mbars was maintained by immersion of the biogas exit tube to 20 mm below the water surface of a bottle that served as a biogas check-valve. Biogas exited the check-valve bottle via a 4 mm tube that carried the biogas to an on-line analyser comprising flow through CH₄ measurement (IR) and a mass flow meter (thermal mass principle) calibrated with a mixture of 60% CH₄ and 40% CO₂ and sampled every second. The biogas entered the gas analyser in pulses corresponding to the biogas bubbles observed in the water trap. Existing lab instruments were used during the pilot study for biogas quantification and composition. It was not possible to make accurate flow rate and methane concentration measurements during the pilot study. The mass flow meter requires a flow rate of at least 500 ml/min to measure accurately. This flow rate was not attained regularly. When there was no biogas flow, the biogas in the methane measuring cell was diluted with ambient air. Since biogas flowed in irregular pulses, only the peak methane concentrations should be considered as valid measurements. Although there are major deficiencies in the technique used during this exploratory study, the approaches used made it possible to quantitatively and to qualitatively assess biogas production.

Quantity of biogas produced

Three methods were used to estimate the methane production.

1. Calculation of total biogas volume flow by integration of the area under the mass flow rate curve. The error of this method is due to variations in the gas composition and to the actual flow rate being less than the minimum flow rate required by the manufacturer of the mass flow meter. The advantage of this method is that variations during the day can be monitored because the gas flow rate is recorded every second.
2. Calculation of methane production from the difference in the higher heating value of the solids contained in the influent and the effluent. The error of this method is due the fact that not all the chemical energy lost is transformed into methane.
3. Calculation of methane production from the difference in the volatile solids content of the solids contained in the influent and the effluent. The error of this method is due to inaccuracy of volatile solids determinations and lack of knowledge the biogas yield from volatile solids.
4. Calculation of methane production from the difference in the COD content of the influent and the effluent. The error of this method is due the fact that not all the COD lost is transformed into methane.

The average daily methane production rate determined using these 4 methods is presented in table.

| Method | Mean CH ₄ production [N] | RSD [%] | Remark |
|-------------|-------------------------------------|---------|--|
| Mass flow | 6 | | Assume 65% CH ₄ content of biogas |
| HHV removal | 20 | 42.5 | 36 MJ/m ³ CH ₄ |
| COD removal | 29 | 28 | Assume 0,35 m ³ CH ₄ /kg COD removed |

| | | | |
|------------|----|----|--|
| VS removal | 17 | 69 | Assume 0,6 m ³ CH ₄ /kg VS removed |
|------------|----|----|--|

Table 6 Rate of methane production

Quality of biogas produced

Biogas components were also measured on 3 different days using a multigas detector that sucked biogas from the head space. The gas detector was designed for security and not for analytical use. The measured values are reported in the table.

| Biogas component | Unit | Measured values |
|------------------|----------|-----------------|
| CH ₄ | % volume | 67, 79, 76 |
| H ₂ | ppm | 4, 12, 3 |
| H ₂ S | ppm | 56, 373, >500 |
| NH ₃ | ppm | 3, 77, 31 |

The methane concentration values measured on-line using the IR flow through cell were regularly 10% volume less than the values measured using the multigas detector. Consequently, the actual methane value is assumed to be at least 5% greater than the reported on-line measurement value.

The methane content of the produced biogas was consistently superior to 60% volume and frequently superior to 70% volume.

Robustness

The biological activity of the anaerobic filter is the result of a biofilm that developed on the torrefied wood chip support. Due to the presence of extracellular polymer support matrices and to microbial consortiums, biofilms are known to be more resistant to chemical and hydraulic variations than suspended growth cultures. The tests of biological robustness of the anaerobic filter bioreactor and the range of values tested are.

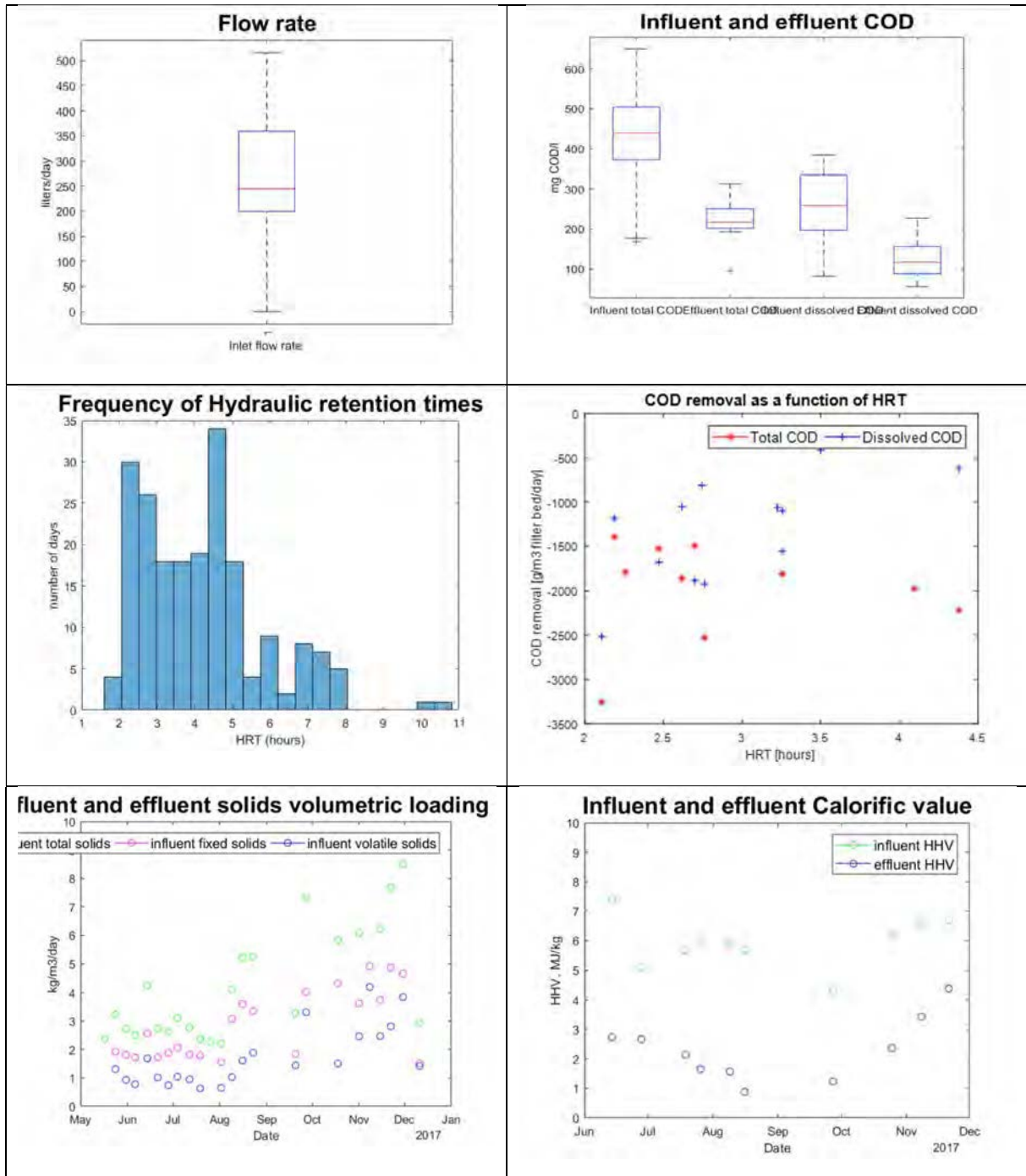
| Variable influent parameter | Range of variation |
|---------------------------------------|--------------------|
| Flow rate [l/day] | 0 to 516 |
| Influent temperature [°C] | 8 to 40 |
| COD load [kg COD/m ³ /day] | 1 to 6 |

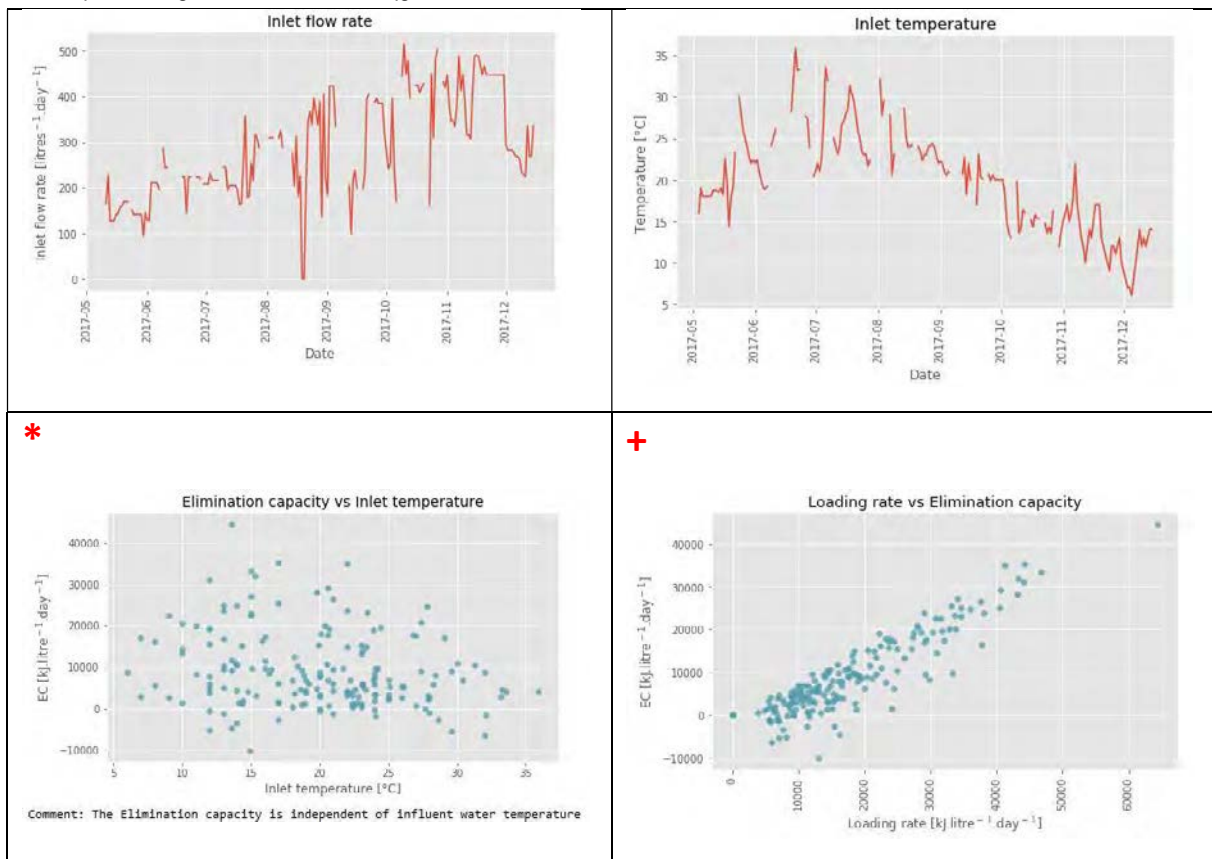
Table 7 Range of important variable influent parameters

The variations in all the measured parameters are presented graphically on request. The general conclusions are that the AFBR is resistant to the tested:

- Influent flow rate variations
- Temperature variations. However, biogas productivity declined as the average temperature declined.
- COD load variations
- Influent nutrient variations

Some figures





* Elimination capacity (EC) = (HHVin – HHVout)/UAF working volume/day.

This result demonstrates that the process was effective at influent temperatures between 12 and 28 $^{\circ}$ C. This is the range of normal sedimentation basin effluent in temperate climates. 90% of the Δ HHV values were obtained using a fitted equation derived from measured temperature and solids calorific value.

+ Loading rate in terms of influent solids calorific value. The observation that the EC increases with loading rate demonstrates that the UAF was not overloaded.

AFBR Residence Time Distribution

Efficient use of the AFBR volume requires dispersion of the influent substances (substrate) throughout the filter bed volume. Interpretation of time series plots of the effluent concentration following pulse injection of a tracer substance gives useful information about dispersion and the effectiveness of a packed bed column like the AFBR. Ideally, the substrate will be dispersed across the entire width of the filter bed and advance as a plug to the effluent side thereby allowing the greatest exposition to the biofilm and the longest reaction time. Comparison of the ideal effluent conductivity curve of an ideal plug flow reactor to the actual conductivity curve gives an indication of the effectiveness of substrate dispersion in the anaerobic filter.

The distribution of substrate in the anaerobic filter was assessed by conducting a residence time distribution study of a concentrated salt solution injected as a pulse immediately upstream of the

anaerobic filter bioreactor². Effluent salt concentration was quantified by conductivity measurements of the AFBR effluent stream. The effluent conductivity profile was compared to an ideal profile. Residence time distribution curves were obtained for six separate injections of salt solutions into the AFBR.

According to chromatography theory, a theoretical plate is described by the time that it takes for solute to equilibrate between a mobile and a stationary phase³ and the number of theoretical plates is a function of the retention time and the width of the effluent peak⁴.

$$N \approx 5,54 \left(\frac{V_r}{W_h} \right)^2$$

N = Number of theoretical plates

V_r = Peak retention time

W_h = Peak width at half peak height

Since efficient use of the anaerobic filter also involves distribution of substrate and contact with a biofilm that is attached to solid packing media, these concepts also apply to an anaerobic filter bioreactor. The data obtained from the experiments was also used to determine the theoretical plate height and optimum mobile phase flow rate.

The range of hydraulic retention times tested was chosen to correspond with the hydraulic retention times tested in the AFBR at the wastewater treatment plant between May and December 2017. After observing that the conductivity profiles corresponded with a well-mixed plug-flow reactor, shorter hydraulic retention times were tested to find the optimum.

The conditions are summarized in the table below and the results are presented in the following tables and graphs.

| | | |
|---|--------|-------------|
| Anaerobic filter bed volume | liters | 42 |
| Total bioreactor liquid volume | liters | 37 |
| Flow rate | l/h | 7,6 to 90,4 |
| Hydraulic retention time (based on total liquid volume) | hours | 0,41 to 4,8 |
| Salt introduction (dissolved in 1.1 liters water) | grams | 360 |

Table 8 RTD study conditions

² Smith, Elliot and James. Characterisation of mixing patterns in an anaerobic digester by means of tracer curve analysis. *Ecological Modelling*, 69 (1993) 267-285.

³ <https://teaching.shu.ac.uk/hwb/chemistry/tutorials/chrom/chrom1.htm>. Accessed April 2018.

⁴ GE column efficiency testing brochure, Application note 28-9372-07 AA.

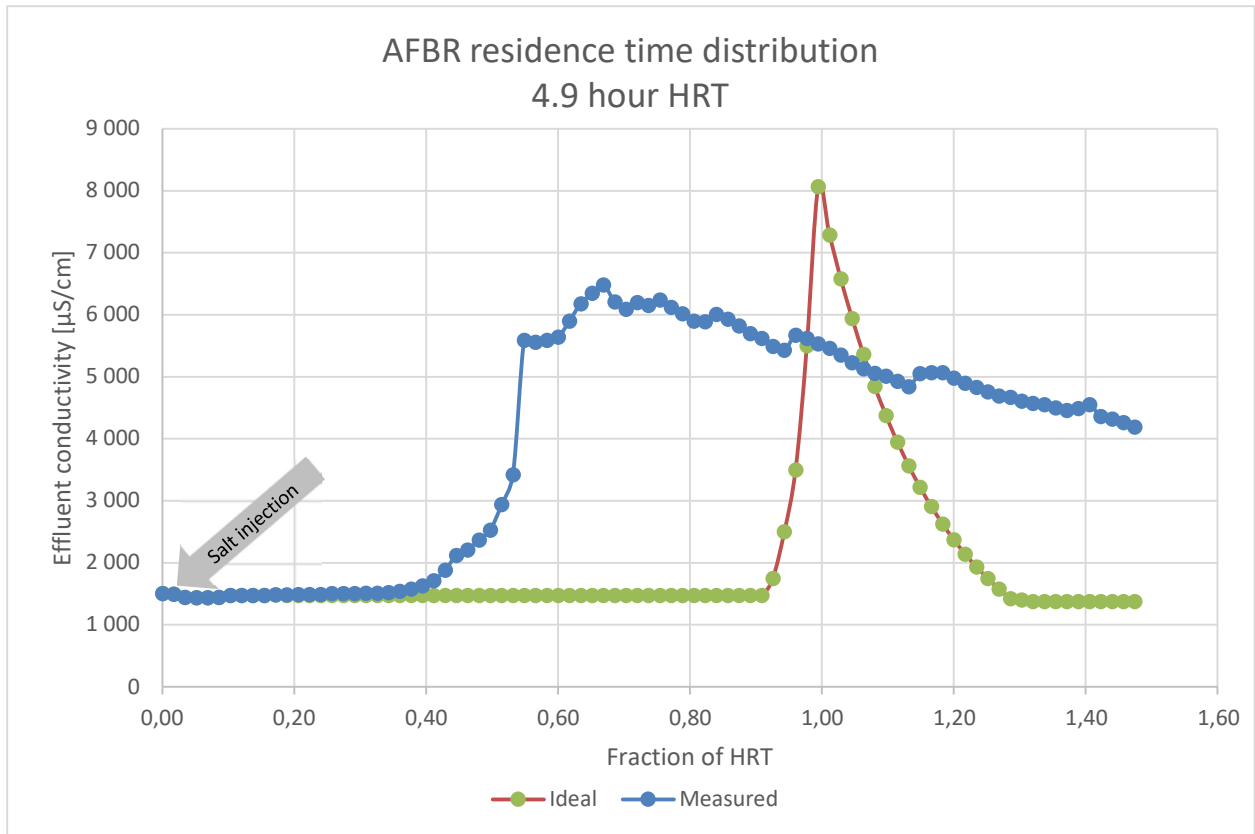


Figure 8 RTD 4.9 h HRT

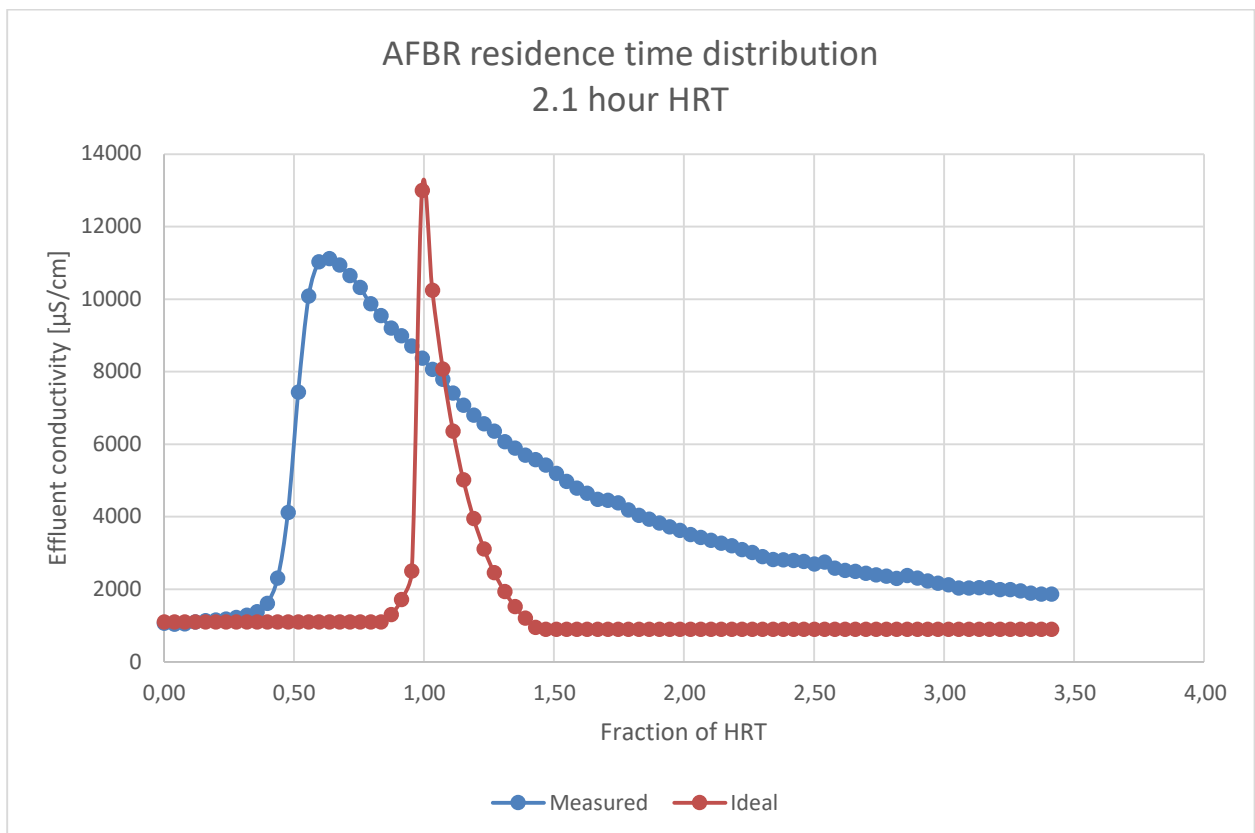


Figure 9 RTD 2.1 h HRT

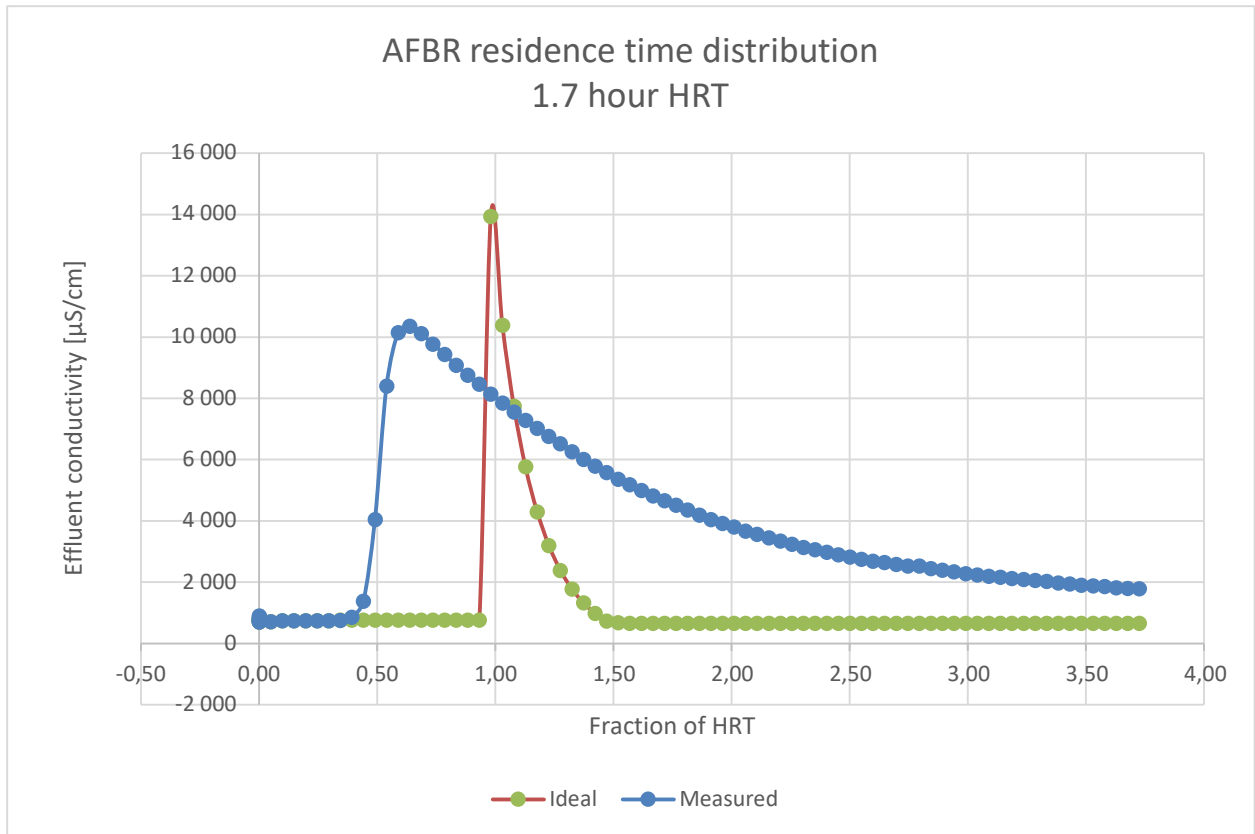


Figure 10 RTD 1,7 h HRT

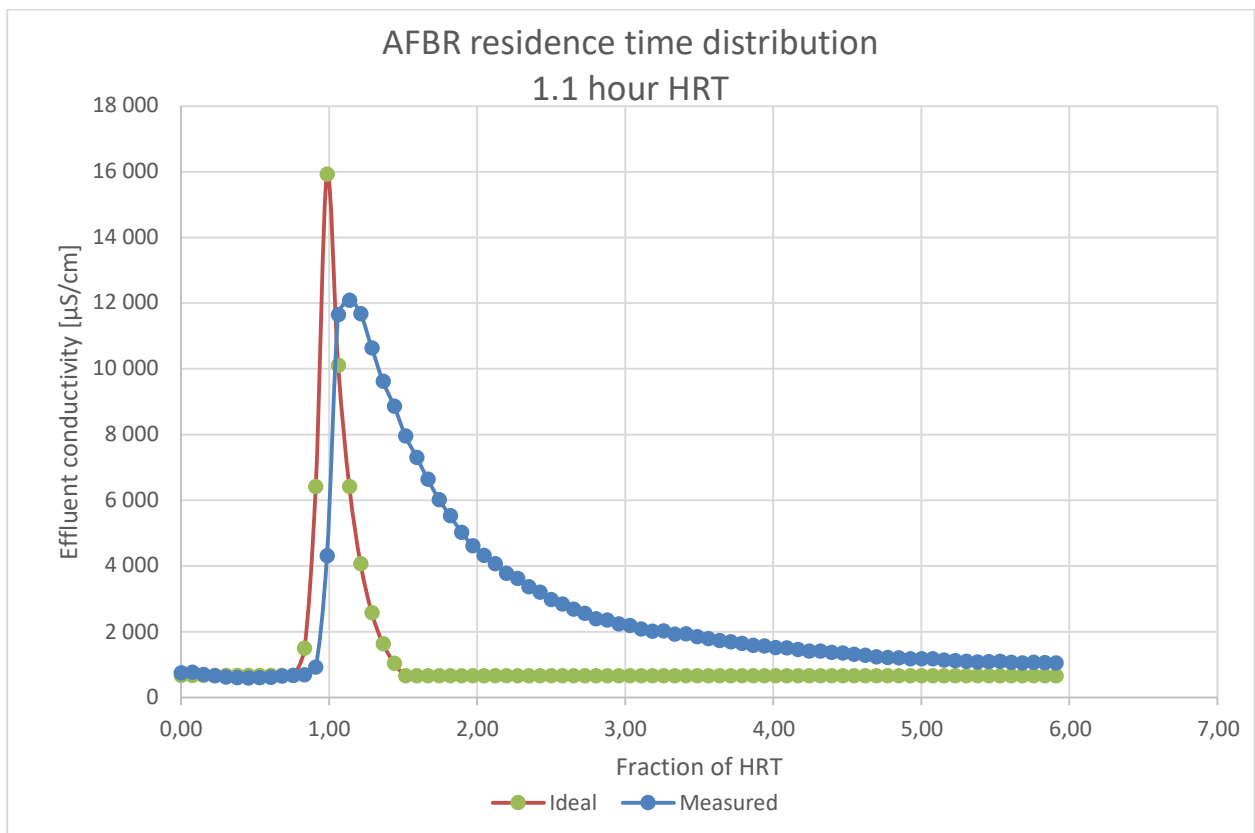


Figure 11 RTD 1,1 h HRT

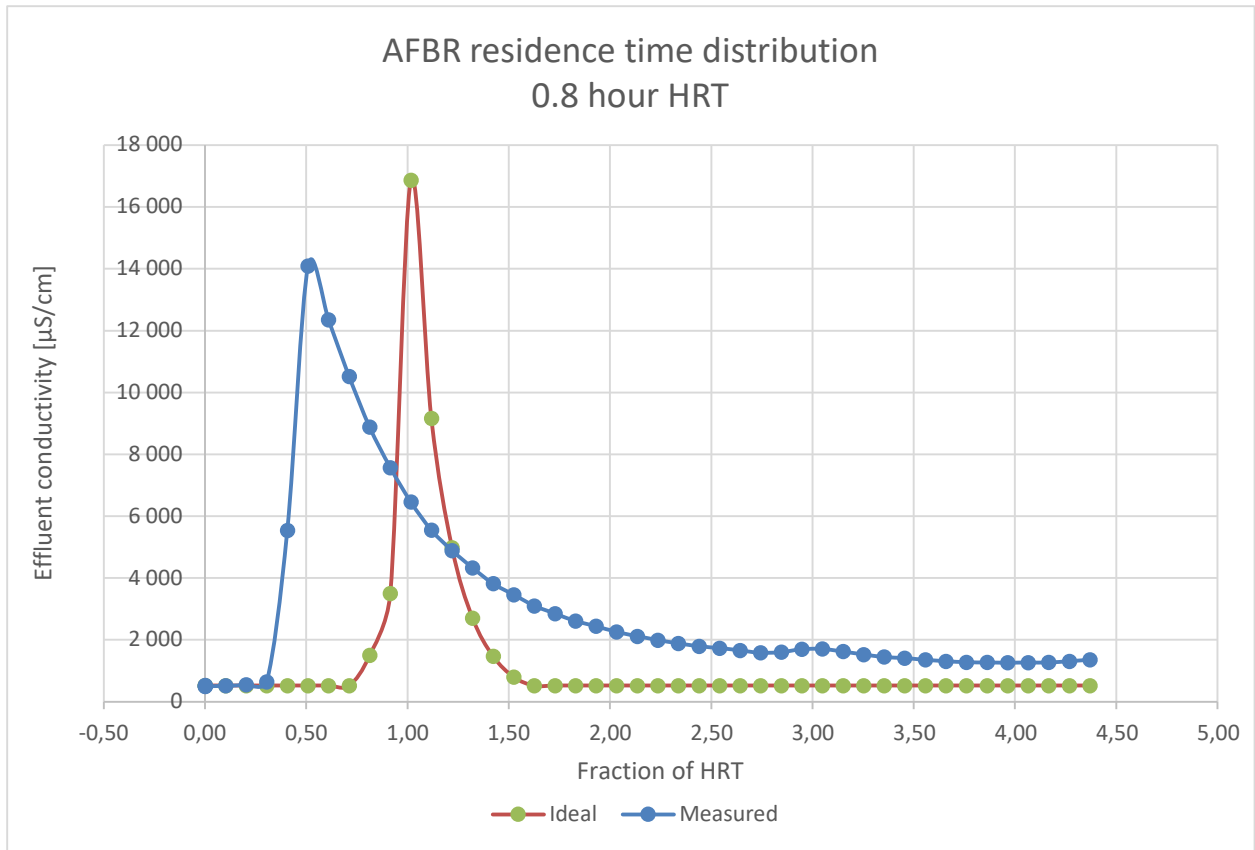


Figure 12 RTD 0,8 h HRT

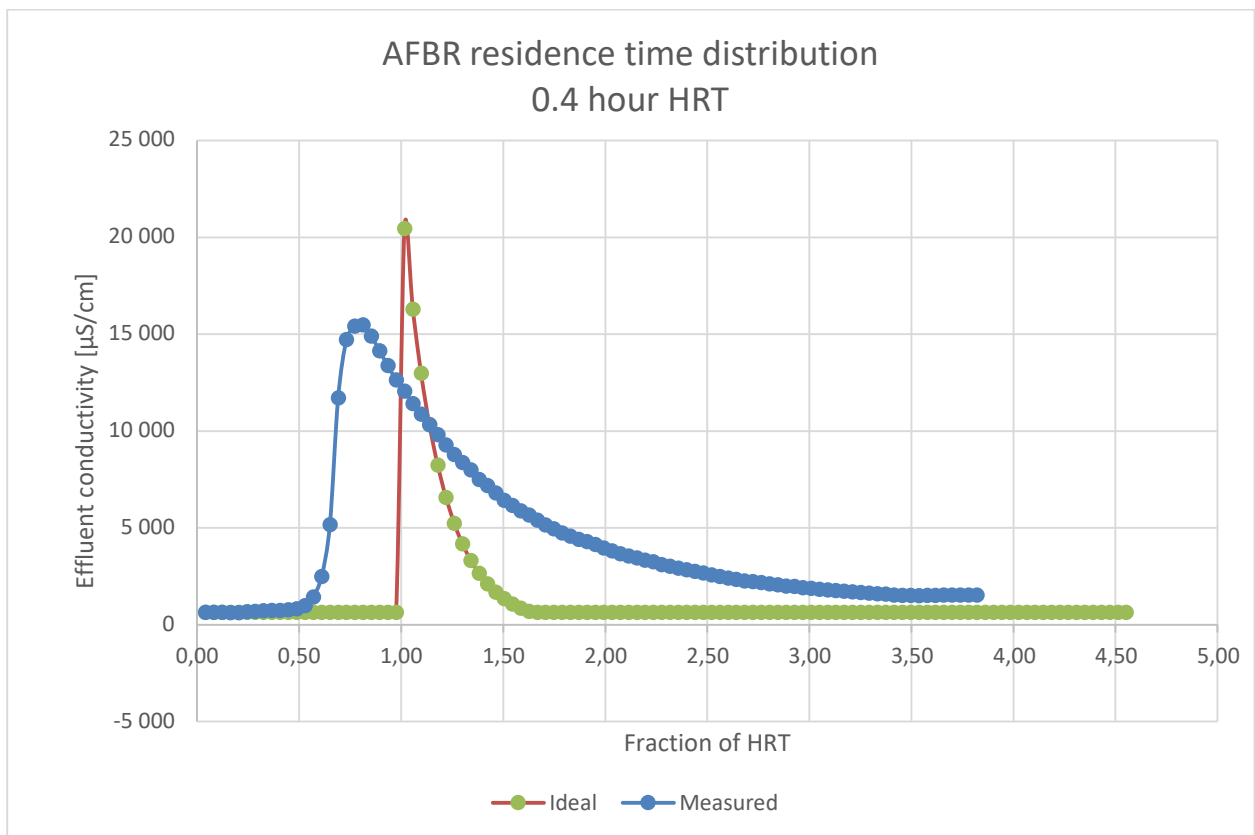


Figure 13 RTD 0,4 h HRT

In the case of a long hydraulic retention time of 4.9 hours, the concentration rises slowly after only 2.5 hours, then abruptly, and then tails off slowly. The slow initial rise indicates short-circuiting. The abrupt rise indicates plug flow behaviour. The appearance of the first peak after only ~ 0.5 HRTs indicates that the effective volume of the experimental AFBR is less than the actual volume. The absence of a narrow peak and the very slow tailing off indicates the presence of dead zones from which salt continues to diffuse during a prolonged period after reaching a peak concentration in the effluent. The conclusion is that the long HRT leads to the creation of short circuits, dead zones, and the inefficient use the anaerobic filter bed volume.

In the case of short hydraulic retention time of 2.1 hours, the concentration rises abruptly after 1.3 hours and then tails off slowly. The abrupt rise indicates plug flow behaviour as expected. The appearance of the first peak after only ~ 0.8 HRTs indicates that the effective volume of the experimental AFBR is less than the actual volume. The slow tailing off indicates the presence of dead zones from which salt continues to diffuse during a prolonged period after reaching a peak concentration in the effluent. The tests at 0,8 and 0,4 hours HRT show similar narrow peaks that arrive before the theoretical HRT. Only the test of a 1.1-hour HRT gave a result where the conductivity peak corresponded with the theoretical HRT. The conclusion is that the short HRT leads to more effective mixing of the substrate and efficient use the bioreactor volume with a 1.1 hour HRT being the best for this particular anaerobic filter. In a future project the hydrodynamics should be optimized, and the HRT reduced to less than 3 hours.

In column chromatography, the height equivalent of a theoretical plate (HETP) is a function of the hydraulic retention time, the RTD curve width and the bed length. The Van Deemter plot is used to compare the HETP to the mobile phase velocity at different hydraulic retention times. The most efficient use of the bed volume is attained at the minimum of the curve.

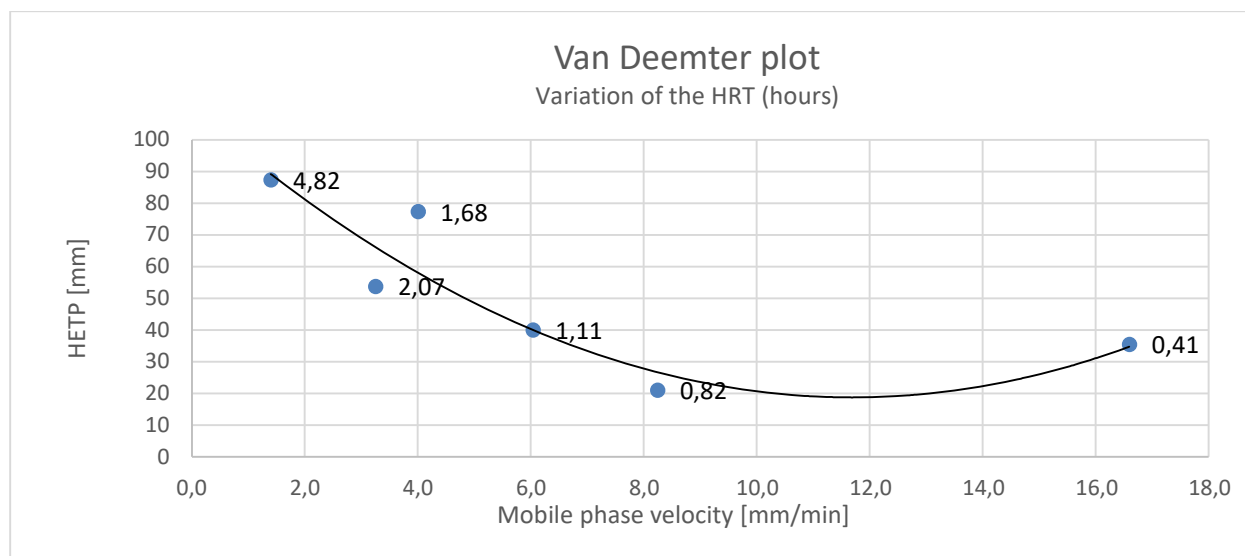


Figure 14 Van Deemter plot

| Tested HRT | hours | 0,41 | 0,82 | 1,11 | 1,68 | 2,07 | 4,82 |
|--|---------|------|------|------|------|------|------|
| Measured flow rate | l/h | 90,4 | 44,9 | 32,9 | 21,8 | 17,7 | 7,6 |
| Peak retention time, t_r | minutes | 49 | 32 | 68 | 38 | 38 | 40 |
| Peak width at half-height, W_h | minutes | 32 | 16 | 47 | 36 | 30 | 41 |
| Number of theoretical plates, N_{tp} | | 13 | 22 | 12 | 6 | 9 | 5 |
| Height equivalent of a theoretical plate, HETP | mm | 35 | 21 | 40 | 77 | 54 | 87 |
| Mobile phase velocity | mm/min | 17 | 8 | 6 | 4 | 3 | 1 |

Figure 15 Data used to make the Van Deemter plot

The recommendations that follow from the residence time study are:

- HRT for the AFBR configuration tested (packing volume, packing dimensions, column geometry, etc.): between 0,4 and 1,1 hours
- Mobile phase velocity: between 6 and 16 mm per minute

Discussion and conclusions

Observed bioreactor performance and projected scale-up

The important average influent and effluent characteristics and the values achieved during the pilot tests are presented in the tables below. Achieved values are those that were maintained stable for several days. Expected values are predicted based on the results of the pilot tests and subjective judgement.

Table 9 Expected influent and effluent characteristics

| Influent | Unit | Overall average | RSD [%] | Achieved | Expected |
|----------------------|--------------------|-----------------|---------|----------|----------|
| Total COD | mg/l | 685 | | | 685 |
| Total solids | mg/l | 646 | 6 | | 650 |
| Volatile solids | % TS | 37 | 22 | | 37 |
| Higher heating value | kWh/m ³ | 1.04 | 36 | | 1 |
| Effluent | | | | | |
| Total COD | mg/l | 347 | 31 | 56 | 150 |
| Total solids | mg/l | 504 | 20 | 400 | 500 |
| Volatile solids | % TS | 29 | 23 | 25 | 25 |
| Higher heating value | kWh/m ³ | 0.29 | 49 | 0.25 | 0.25 |

Table 10 Summary of AFBR performance

| AFBR performance | Units | Overall average | RSD [%] | Achieved | Expected |
|--------------------------------|---------------------------|-----------------|---------|----------|----------|
| Hydraulic retention time | hours | 4.7 | 122 | 2.8 | 3 |
| Temperature | °C | 23 | 30 | 16 | >16 |
| Methane production | vv/day | 0.44 | 42.5 | 0.6 | 0.7 |
| Biogas CH ₄ content | % volume | 65 | 15 | 70 | 65 |
| Total solids loading rate | Kg TS/m ³ /day | 4.6 | 49 | 7 | 7 |
| Organic loading rate | kg VS/m ³ /day | 1.7 | 61 | 4 | 10 |
| COD reduction | % | 58 | 28 | | 75 |
| HHV reduction | % | 71 | 14 | 85 | 80 |

The average inlet COD (429 mg/l) x average flow rate (268 l/day)/ filter bed working volume (56 litres) = 2,1 kg COD/m³/day.

Table 11 AFBR description of industrial scale AFBR

| Anaerobic filter bioreactor description | | Achieved | Expected |
|---|--|-----------------|----------------------------|
| Biofilm specific area | m ² /m ³ bulk volume | 360 | 360 |
| Biofilm thickness ⁺ | µm | 100* | 300 |
| Biofilm specific mass | mg VSS/l reactor | 36'507 | 164'280 |
| Effective volume* | % of actual volume | 67* | 99 |
| Bed volume requirement | m ³ /MLD influent | 196 | 29 |
| Biofilm support material | | Torrefied chips | Torrefied chips or plastic |

⁺ assumed mean thickness based on visual observation that the chip surface was not yet completely covered.

* Based on tracer studies that show that for the average HRT of 4,7 hours observed during the study, the peak effluent salt tracer concentration occurs at 67% of the theoretical HRT.

The main needs for further development are:

- Hydrodynamic models and detailed design of the AFBR bioreactor
- Design, construction and testing of features that will be required at the industrial scale (e.g. high-rate primary sedimentation basin, back-washing, bed materials, heating mode, effluent solid-liquid separation)
- Development of biochemical and kinetic model
- Large-scale (> 4 m³ bioreactor) long duration (18 months) testing
- Design, construction and testing of a thermal energy collection, storage and distribution system

The AFBR project can be conducted as a standalone project or, preferably, as part of a project to develop a completely new wastewater treatment plant incorporating a new primary treatment process, nutrient removal, algae photo bioreactors, CO₂ recycling, advanced biogas cleaning techniques and combined heat and power concepts, novel final effluent polishing processes, and water reuse.

Biofilms and packing material

The results of the three different methods used to assess the biofilms lead to the following main conclusions:

- A robust biofilm developed on the torrefied wood chip supports
- Torrefied wood is an acceptable biofilm support material
- A biofilm developed on the plastic supports
- Biofilm development is slow. It appears that at least 1 month is required for establishment of the biofilm and that one can expect the biofilm to continue to increase in surface coverage and biomass over time
- The biofilm tolerated cold conditions (8°C). Maintenance of at least 16°C inside the biofilter bed appears to be a reasonable objective for future development

Further evidence of the existence and activity of a biofilm is given by the measured biochemical conversions. In particular, the short hydraulic retention time, the higher heating value reductions and the COD removals would not have been possible without active immobilized biomass.

The observation that the UAF elimination capacity increased during the study as the hydraulic retention time decreased and as the temperature decreased suggests that the mass of the biofilm was increasing during the study and had not reached a limit. Further increasing the mass of the biofilm might result in even better performance.

Due to the strict limits on effluent turbidity and the unknown effect of torrefied wood particles on conventional wastewater treatment processes, further work will be focussed on the use of synthetic biofilm support materials. Differences between the type of plastic material were observed in terms of biofilm coverage and in specific biofilm activity as determined by the ADH assay. The micropollutant adsorption capacity of torrefied chips should be investigated. Due to their low cost compared to plastic, torrefied chips might be appropriate biofilm support materials in lower income countries.

Acknowledgements

This research project was financially supported by the Swiss Innovation Agency Innosuisse and was part of the Swiss Competence Center for Energy Research SCCER BIOSWEET. A student intern made a major contribution to adapting the DHA activity assay and to sampling and analysis of the physical-chemical parameters.



Schweizerische Eidgenossenschaft

Confédération suisse

Confederazione Svizzera

Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Mobile environmental bioprocess data acquisition unit for artificial neural network model development

Application

Number:

101.992 IP-EE

Title in English:

Mobile environmental bioprocess data acquisition unit for artificial neural network model development

Project Duration:

Start: 01.05.2022

End: 30.04.2023

Duration: (12M)

Requested Innosuisse Funding incl. Overhead

CHF 311,459.35

Special Funding Measure

No special funding measure

Research Partner(s)

UNIL - Université de Lausanne

Implementation Partner(s)

1. Introduction

Abstract

Please note, the abstract will be published in Aramis

Design, construction, and operation of a mobile bioreactor test system for the acquisition of data required to build predictive models using artificial neural networks. The project results will be used to design a complete system to convert wastewater to valuable products.

Management Summary

Please note, the executive summary will not be published in Aramis

According to UN Water, industrial water consumption is responsible for 19% of global water use and around 380 billion m³/year of water can be recovered (2021, UNESCO, WWAP). Many wastewater streams are dilute aqueous suspensions or solutions such as municipal wastewater and industrial streams from the dairy, beverage, prepared foods, meat processing, mining, and solid waste management industries. We will develop novel bioreactors for lactic acid production and for COD reduction using the wastewater streams from cheese dairies. To do this, we will build and operate a mobile data acquisition system that has been designed to acquire a large and dense dataset as required to build Artificial Neural Network predictive models. In contrast to the conventional factorial designed experiments and regression analysis approaches, the proposed workflow features efficient data acquisition and simulation using Artificial Neural Network models to identify the important parameters and to set mechanical features and operational parameters of the novel treatment system to their optimal values. By working with real industrial streams and using large bioreactors, the results from tests using the mobile data acquisition unit are expected to be more useful to design engineers than results obtained from laboratory experiments. Additionally, this workflow makes it possible to tailor the bioreactors to match site conditions thereby achieving high performance in terms of Capex, Opex, pollution removal and stability and constitutes a unique selling proposition. The project deliverables include construction and operation of a mobile data acquisition unit, experimental results, modelling, and simulation as required to design a complete industrial system. Commercial implementation requires completion of field data acquisition and design of the first commercial unit. The start-up to be created during the project aims to have revenues from sales within 12 months of project completion.

Context

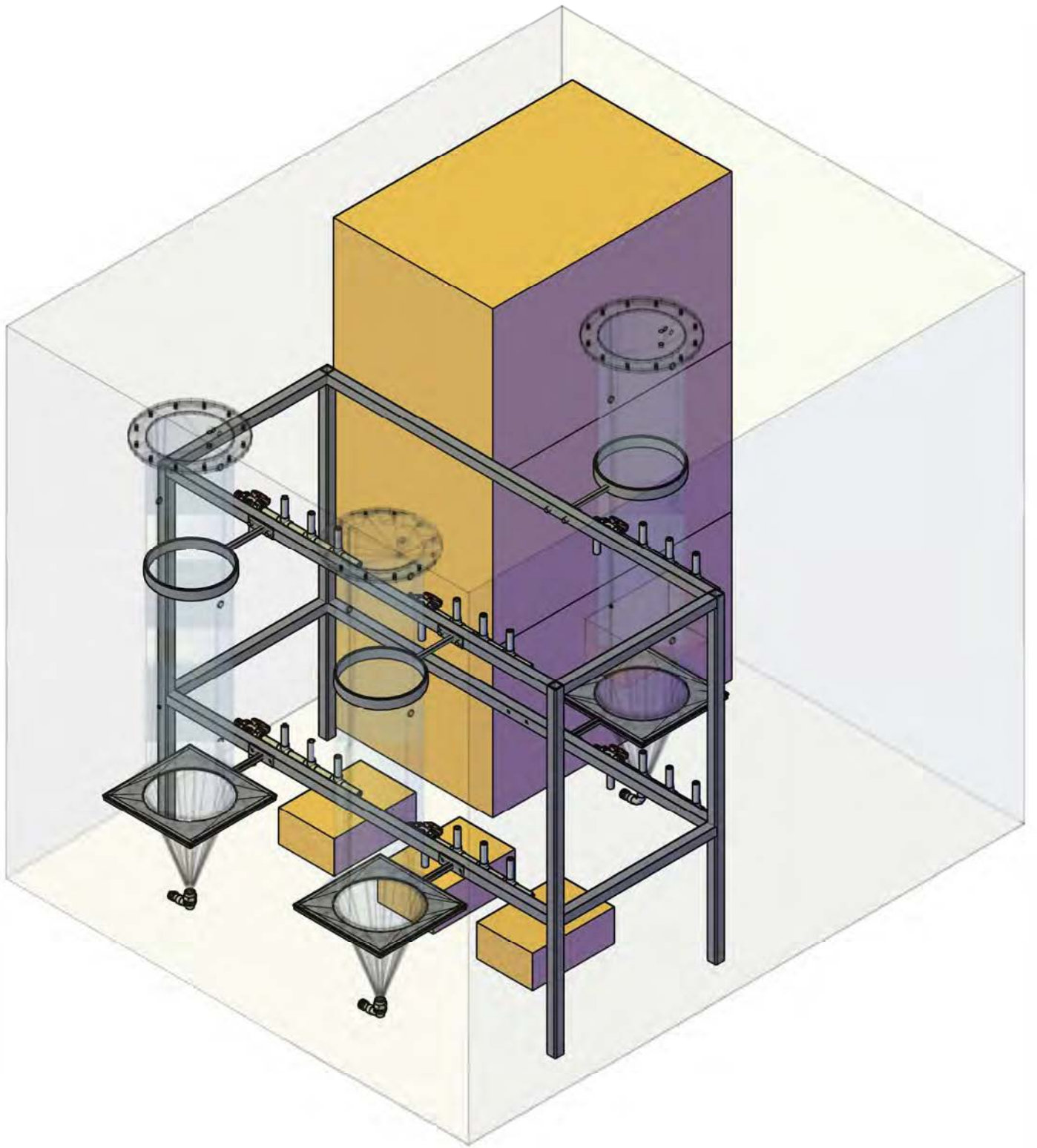
| | |
|--------------------------|--------------------------------------|
| Thematic Area: | Energy & environment |
| Cluster: | Environmental technologies/cleantech |
| Special Funding Measure: | No special funding measure |

Is the application a follow-up of an idea financed in the frame of an Innosuisse NTN Booster?

No idea out of an NTN Innovation Booster

Have any of the topics in this application been previously developed with Innosuisse or other funding instruments?

Innovation cheque: 47096.1 INNO-EE Data acquisition unit for deep neural network model development



Mobile Data Acquisition System
 Container volume: 15.7 m³
 Auxiliary equipment: 3 x reservoirs,
 3 x pumps, 1 x electronics cabinet

EBR_system

2021-04-21

Mark McCormick

1:20

Unil

47096-1
 INNO-EE

Machine learning to improve environmental bioprocess design

Mark McCormick

What is the problem?

Bioprocesses can be defined as controlled conditions that result in the biological agent mediated transformation of a substrate chemical into a specific product chemical.

In practice, bioprocesses are embedded in complex systems comprising physical, chemical, and biological parameters with non-linear relations. Consequently, they are difficult to model, design and control.

What is needed are methods for system identification, simulation and process optimization.



Source: pexels-tom-fisk



Source: pexels-tom-fisk

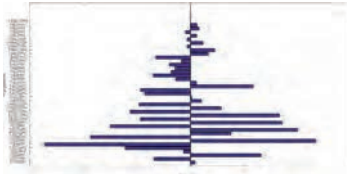


Source: pexels-skitterphoto

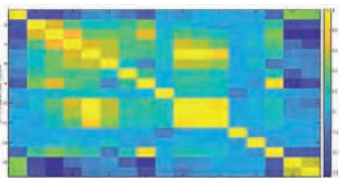
Why use machine learning?



Use of Hotelling's T² statistic to identify periods of process instability



Evaluation of the contribution of process parameters to process instability



Cross correlation of process parameter values

The use of machine learning and neural network environmental process data analysis methods helps to gain insights and understanding. For example:

Non-linear system identification

Latent variable models

Predictive models

Time series analysis

Process monitoring

Trouble shooting and root cause analysis

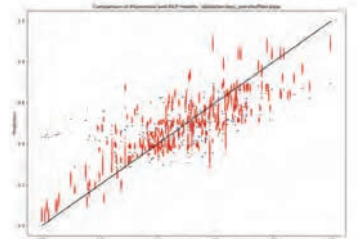
Optimisation

Simulation

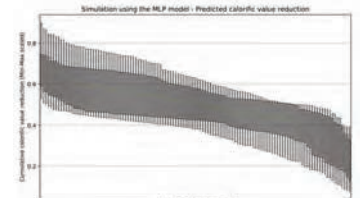
Selection of mechanical features

Analysis of very large datasets

Decision trees



Numerical predictive models of a fixed bed bioreactor
Predicted versus True values
• Red: Multilayer Perceptron neural network
• Blue: 4th degree polynomial



Simulation of different constructive feature values in a fixed bed bioreactor using a Multilayer Perceptron neural network model

Perspectives

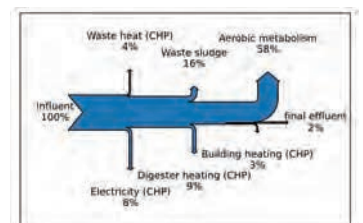
Integrating ecosystems, wastewater treatment and energy production to improve environmental quality

Problem

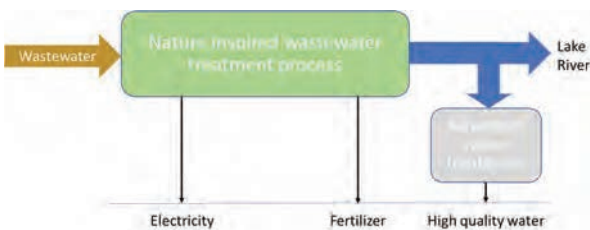
According to UN Water, 2.5 billion people do not have access to improved sanitation. Billions more are served by underperforming wastewater treatment systems. The conventional activated sludge process generates significant CO₂ emissions due to bacterial respiration and electricity consumption.

Current requirements include:

- Economic affordability
- High energy efficiency
- Carbon, nitrogen and phosphorous recycling
- Reduction in GHG emissions
- Resource recovery



Flow of potential energy (NCV) through a conventional activated sludge wastewater treatment plant



Solution

A new wastewater treatment plant concept inspired by natural processes and systems is possible. The use of machine learning methods in the design process and in optimization would contribute to effectiveness. Compared to a conventional activated sludge plant, some advantages of the new plant include:

- 90% reduction in electricity consumption
- 70% reduction in CO₂ emissions
- 80% reduction in operating costs

Mark McCormick
Environmental engineer, M.Sc.
PhD candidate in information systems
Unil, HEC, DESI
mark.mccormick@unil.ch
Tél: +41 78 604 52 42



Conclusions and perspectives

This work has demonstrated the use of machine learning methods to understand and to model 3 large datasets obtained from very different experiments in the field of environmental engineering. In all 3 experiments, the use of machine learning methods made it possible to achieve more informative numerical results and visualizations and more powerful models than could have been achieved using conventional statistical methods and mechanistic models.

The work also contributed to the development of a new workflow that eliminates some conventional practices used in environmental bioprocess development and consequently is expected to lead to significant savings in development time and costs and to the development of bio processes that are more effective and less expensive to build and operate. Among the conventional practices that were changed or eliminated are:

- The use of the conventional factorial design of experiments and regression analysis approach
- The use of laboratory studies under highly controlled conditions that often are very different from industrial and environmental conditions
- Labor intensive off-line analysis of samples
- The use of inconvenient stoichiometric and kinetic based models such as ADM-1
- The use of mechanistic models that do not accurately account for the non-linearity and complexity of biological processes

These conventional practices will be replaced by a new development framework and workflow that includes:

- Acquisition of data during short-duration studies using medium sized bioreactors (200 liters) and real waste streams under real industrial conditions
- A black-box Artificial Neural Network (ANN) model that ignores the biochemical mechanisms and kinetics
- Determination of the design specification of the industrial-scale bioreactor by simulation using the ANN model

The goal of this thesis was practical. The goal was neither to contribute to the theoretical understanding of neural networks nor to develop new methods in the field of machine learning. My personal objective was to learn to apply machine learning to the kind of datasets that I acquire during my work as an environmental engineer. This goal was achieved as demonstrated by the progressive acquisition of machine learning skills required to produce 3 articles. Among the specific skills acquired were, programming in Python, use of the Bash shell to access the High-Performance Computing (HPC) clusters at the University of Lausanne, use of the Scikit-Learn, Matplotlib and Pandas libraries for data pre-processing and visualization, the use of open source libraries for

statistical data analysis, and use of the Keras API to build Artificial Neural Networks based on the MLP, LSTM and CNN architectures.

The use of PCA, PLS regression, lag-time analysis, control charts based on Hotelling's t-squared statistic were already well known to the field of process control. The contribution of this work was to demonstrate a new application of these techniques to the biomass torrefaction process.

The use of a state image approach to monitoring and controlling a fed batch anaerobic digestion process was demonstrated. The 1-D CNN proved to be an effective architecture for this task involving recognizing the chemical state conducive to feeding the digester.

A MLP DNN was built to model a continuous, high-rate bioprocess conducted in a fixed bed anaerobic digester. The model was saved and reloaded to conduct simulations using new data. The workflow will be used to determine the mechanical design parameters of industrial bioreactors for wastewater treatment.

While working on my thesis, I designed a mobile data acquisition system (Innocheque project n° 47096-1-INNO-EE). The system can be used to acquire a very large dataset as required to build artificial neural network models. The application method involves temporarily installing the mobile unit at the study site and operating 3 bioreactors according to an experimental plan that covers the range of plausible industrial operating conditions. The results are intended for use in designing wastewater treatment and conversion equipment that can be used to add value to wastewater. With this goal in mind, contacts were established with potential partners to develop projects in the areas of lactic acid production from whey, hydrogen production, and municipal wastewater treatment. An application for Innosuisse funding to build and operate the mobile data acquisition (Innovation project n° 101.992 IP-EE) was rejected in April 2022.

The design of equipment and control strategies for municipal wastewater treatment plants is a major potential application of machine learning and artificial neural networks. According to UN water, more than 2.5 billion people have no wastewater treatment service. Billions more are served by inadequate or poorly functioning wastewater treatment systems. Among the root causes of the problem are high construction and operating costs and electricity consumption. An inexpensive, robust, and energy positive wastewater treatment system might help to solve this problem. Contacts have been established with the goal of delivering a preliminary design report including a process flow diagram, material balance and site layout adapted to the local partner's conditions. The long-term goal of this work is to design and build wastewater treatment plants each serving between 50'000 and 500'000 population equivalents. Since bioprocesses, including upflow anaerobic filters, are major components of the proposed system, the knowledge about machine learning and artificial neural networks that I gained while preparing this thesis will be applied to this project.

Although I have not yet applied machine learning and artificial neural networks to ecosystems, I aim to collaborate on projects focused on ecosystem services such as constructed wetlands. Wetland ecosystems involve complex interactions between system features such as land use, biotic and hydrologic conditions, soil type and water physical-chemical quality. These interactions have a

time and space dimension. Artificial Neural Networks (ANN) have demonstrated their capacity to map complex and non-linear relations between data. Consequently, in the study of ecosystems the ANN approach to modeling is expected to make it possible to develop accurate predictive models where prediction accuracy is evaluated in terms of the error between predicted and true values. Machine learning and Neural Network approaches have proven to be very useful tools for identifying latent variables and for mapping relations between parameters that measure apparently very different features such as those that might be discovered in ecosystems.

In this context, the objective is to provide data-based answers to questions asked about the studied ecosystems. One research question is: what is the most appropriate workflow as defined by data collection, data preprocessing, model architecture, training, and deployment? Work towards the conservation and restoration of wetlands would benefit from numerical models of the biological and physical processes occurring within the studied ecosystems. These models would have many uses including assessing the role of different biodiversity, hydrologic and physical-chemical dimensions, and their relative importance to the studied ecosystems. The models could also be used to identify indicators and targets for use in evaluating conservation and restoration actions. Numerical models can also be developed with the goal of simulation and prediction of future states of the studied areas. In general, having the capacity to make simulations and predictions would benefit planning and upscaling of conservation and restoration work. A poster on this topic was presented in June 2022 at the forum “Les Rencontres de l’Eau” hosted by the University of Lausanne.

This thesis demonstrated applications of machine learning and artificial neural networks to the analysis and modeling of environmental processes. It would not have been possible for me to obtain the results using the numerical and statistical methods that I knew before my doctoral studies. The work revealed latent structures in the raw data and generated useful information about the underlying processes. The information has practical use in the development of process control strategies and in the specification of mechanical constructive features of the equipment.

Choosing to study machine learning and artificial neural networks was a good decision because I gained a better perspective on data science and information systems and the acquired skills have many practical applications. After the thesis, I will continue to improve my skills in machine learning and use these skills to solve problems in the field of environmental engineering.