Multiple target data-driven models to enable sustainable process manufacturing: An industrial bioprocess case study

Oliver J. Fisher [a], Nicholas J. Watson [a], Laura Porcu [b, c], Darren Bacon [b], Martin Rigley [b], Rachel L. Gomes [a, *]

[a] Food Water Waste Research Group, Faculty of Engineering, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom
[b] Lindhurst Engineering Ltd., Midland Road, Sutton in Ashfield, Nottinghamshire, NG17 5GS, United Kingdom
[c] Energy Innovation & Collaboration, University of Nottingham, Jubilee Campus, Nottingham, NG8 1BB, United Kingdom

ABSTRACT

Process manufacturing industries constantly strive to make their processes increasingly sustainable from an environmental and economic perspective. A manufacturing system model is a powerful tool to holistically evaluate various manufacturing configurations to determine the most sustainable one. Previously models of process manufacturing systems are typically single target models, trained to fit and/or predict data for a single output variable. However, process manufacturing systems produce a variety of outputs with multiple, sometimes contradictory, sustainability implications. These systems require multiple target models to find the most sustainable manufacturing configuration which considers all outputs. A novel bioprocess that treats process wastewaters to reduce pollutant load for reuse, while simultaneously generating energy in the form of biogas was studied. Multiple target models were developed to predict the percentage removal of chemical oxygen demand and total suspended solids, in addition to the biogas (as volume of methane) produced. Predictions from the models were able to reduce wastewater treatment costs by 17.0%. Eight models were developed and statistically evaluated by the coefficient of determination (R^2), normalised root mean square error (nRMSE) and mean absolute percentage error (MAPE). An artificial neural network model built following the ensemble of regressor chains demonstrated the best multi target model performance, averaged across all the bioprocess's outputs (R^2 of 0.99, nRMSE of 0.02, MAPE of 1.74). The model is able to react to new regulations and legislation and/or variations in company, sector, world circumstances to provide the most up to date sustainable manufacturing configuration.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Enhancing the sustainability of process manufacturing systems has gained growing attention over the last few decades (Te Liew et al., 2014; Al-Sharrah et al., 2010; Bakshi and Fiksel, 2003). For a manufacturing system to be considered truly sustainable it needs to maximise environmental and social benefits, in addition to economic benefits (Stock and Seliger, 2016). This is particularly important for process manufacturing, which have multiple outputs, each with several and sometimes contradictory, sustainability implications. Process manufacturing is the thermal and/or bio/chemical conversion of resources to products, co-products, by-products and wastes streams (Fisher et al., 2018). Process manufacturing covers a range of industries, including chemicals, food and drink, paint, polymer, pharmaceutical, steel, and fast-moving consumer goods (FMCG). All the process outputs (co-/products, by-products and wastestream composition/volume) require consideration to achieve sustainable manufacturing. This paper investigates how multiple target modelling of a process manufacturing system will provide manufacturers with the tools to holistically evaluate the sustainability of a manufacturing system.[1]

---

[1] Artificial neural network: ANN; Chemical oxygen demand: COD; Coefficient of determination: R^2; Data-driven model: DDM; Dissolved Oxygen: DO; Ensemble of regressor chains: ERC; Maximum correlation chain: MCC; Mean absolute percentage error: MAPE; Methane: CH_4; Normalised root mean square error: nRMSE; Radom forest: RF; Regression chain: RC; Support vector regression: SVR; Total suspended solids: TSS.

Modelling a manufacturing system can aid decision making to identify the most sustainable manufacturing configuration (e.g. feedstock type and source, process conditions' set point, product specification) that may have previously been unknown (Fisher et al., 2018). Furthermore, the model may react to new regulations and legislation and/or variations in company, sector, world circumstances that alter what is termed sustainable, be that economically or environmentally, to provide the most up to date sustainable manufacturing configuration. A powerful tool to achieve this are data-driven models (DDMs). Data-driven models are developed using algorithms to fit and analyse the data and may also be used to make predictions (Fisher et al., 2020). Data-driven models are able to model complex nonlinear relationships within multivariate data (Wang et al., 2018). However, the accuracy of an DDM's predictions relies on the modelling data being representative of the manufacturing system (Coley et al., 2018).

The majority of published research utilising DDMs within process manufacturing systems are single target models. A single target model is trained to fit or both fit and predict data with a single, categorical or numerical target variable of a manufacturing system (Kim, 2017). Recent uses of single target models in process manufacturing show their capability to model: adsorption (Hamid et al., 2016), clean-in-place systems (Escrig et al., 2019), fluid mechanics (Ling et al., 2016), heat transfer (Baghban et al., 2019), mass transport (Kolouri et al., 2017), reaction networks (Ulissi et al., 2017) and synthesise design (Coley et al., 2018). These models may be utilised to discover the most sustainable configuration of the various manufacturing processes. For example, using DDMs to increase removal rates of heavy metals from industrial wastewaters (Halder et al., 2015). Conventional metal removal techniques have negative sustainability implications (high operational cost, incomplete removal, and generation of toxic residuals) (Shahin et al., 2019). By using response surface methodology and artificial neural networks, Halder et al. were able to predict the process conditions that would maximise the chromium (VI) removal efficiency by superheated steam activated granular carbon prepared from non-useable coconut shell, a sustainable alternative to conventional metal removal techniques (Halder et al., 2015). However, this work did not consider that industrial wastewaters often contain a variety of pollutants including heavy metals (e.g. cadmium, chromium, cobalt, copper, nickel and mercury) (Shahin et al., 2019). If instead a multiple target model was developed, it would enable process conditions to be determined that would maximise the removal of multiple heavy metals and/or incorporate economic outputs into the model, increasing the potential of the Halder et al. process as a sustainable alternative for the treatment of heavy metals in industrial wastewaters.

Examples exist of multiple independent single target models, which are combined to fit and/or predict multiple outputs of a process manufacturing system. For example, Lesnik and Liu modelled the power density, chemical oxygen demand (COD) removal, and Columbia efficiency of a microbial fuel cell used to treat wastewater (Larson Lesnik and Liu, 2017). The authors achieved this by developing three single target models (Larson Lesnik and Liu, 2017). However, the authors did not explore developing a single model capable of predicting all three outputs. If they had, the authors could have evaluated how the microbial fuel cells process variables could have been manipulated to maximise power density, COD removal and Columbia efficiency simultaneously. Furthermore, building multiple single target models of a system assumes each output is independent of the other(s) and thus fails to capture any relationship between the multiple outputs and require more computational power (Melki et al., 2017). An alternative approach is to build a model that integrates the multiple outputs into one multiple target model that captures the relationships between the output variables and requires less computational power (Borchani et al., 2015). The development of multiple target models has proven successful when applied to a wide range of practices: bioinformatics (Liu et al., 2010), chemometrics (Burnham et al., 1999), ecology (Kocev et al., 2009), gene function prediction (Kocev and Ceci, 2015), natural language processing (Jeong and Lee, 2009), stock price forecasting (Xiong et al., 2014). There are a limited number of examples of multiple target models built for process manufacturing systems (Curteanu et al., 2011; Liu et al., 2017). For example, Curteanu et al. developed a multiple target artificial neural network to predict the removal of chlorophyll, COD and total suspended solids (TSS) by electrolysis process in wastewater treatment (Curteanu et al., 2011). Although, the authors did not demonstrate how the multiple target model may be used to gain insight or improve the manufacturing system. It is important to not only demonstrate how a multiple target model may be successfully developed but also the methods by which these models may evaluate changes to the manufacturing system (e.g., changes in feedstock characteristics, supply chain, process set points, product quality) in order to find the most sustainable manufacturing configurations. There is a growing body of work (Zarte et al., 2019; Saad et al., 2019; Min et al., 2019) that aims to provide process manufacturers with the decision support systems necessary to increase their sustainability. Multiple target models may act as a tool to improve decision making by enabling these systems to consider all outputs concurrently in one model.

This work develops a multiple target model of an industrial bioprocess that treats manufacturer's wastewaters to improve water quality and reuse, whilst simultaneously generating bioenergy. A multiple target model was necessary to investigate how the different process conditions affect both the pollutant removal rates from the wastewater, and the volume of biogas generated. This model is then utilised to find the most sustainable process conditions for multiple manufacturing environments (various feedstock compositions and effluent disposal costs). More information about the case study is given in Section 1.1. Due to the tendency to model process manufacturing systems using only single target modelling techniques, an overview of the multiple target modelling techniques used in this article has been included in Section 2.

1.1. Bioprocess (H^2AD) case study for waste valorisation

Manufacturers are under pressure to ensure and increase sustainability throughout their systems and an intelligent resource use strategy is paramount to achieving this. Two key resources that process manufacturers consume are water and energy. Water demand in the manufacturing sector is expected to rise by 400% between 2012 and 2050 (OECD, 2012), while energy consumption is projected to increase by 1.2% per year until 2040 (U.S. Energy Information Administration, 2016). Process manufacturers are in an advantageous position to reduce demand for water and energy because their systems offer an opportunity to recover both from waste streams via waste valorisation technologies (Fisher et al., 2018). Waste valorisation or the circular economy refers to industrial processing activities aimed at reusing, recycling, or recovering resources from waste (Kabongo et al., 2013). In the UK, Lindhurst Engineering Ltd. in partnership with the University of Nottingham has developed a technology called "H^2AD" (H2AD, 2020). The H^2AD process combines bioelectrochemical systems (BES) and anaerobic digestion technologies. Bioelectrochemical systems are capable of converting chemical energy into electrical energy (and vice-versa) by employing microbes as catalysts (Bajracharya et al., 2016). The H^2AD is also capable of treating a variety of wastewaters, from sources including agriculture, brewing, soft drinks, foods and bio-

manufacture residues, to reduce the pollutant load and improve the water quality for reuse, whilst simultaneously generating bioenergy. The H^{2}AD is targeted at treating the wastewaters from SME manufacturing processes due to its modular, low-cost design.

The H^{2}AD system was selected to demonstrate the techniques developed in this paper because the bioprocess has multiple objectives to optimise (multi-pollutant remediation from water and biogas generation). These two objectives come into conflict with one another when deciding on which process conditions to use when operating the H^{2}AD. For example, increasing the hydraulic residence time (HRT) of the bioprocess increases the time available for the system to reduce pollutants in the wastewater. However, this reduces the rate of addition of wastewater to the system, reducing the fuel available for biogas generation. This is a simplistic example, the actual relation between HRT and the objectives are more complex (due to the fact the bioprocess is a complex non-linear living system), hence the need for a DDM of the system.

The H^{2}AD system has two objectives: 1) improve the wastewater quality for reuse and 2) energy generation from waste streams. This was represented in the model by 3 outputs: percentage of COD removal (%COD removal), percentage of total suspended solids (TSS) removal (%TSS removal) and daily volume of methane produced by the H^{2}AD bioprocess. The pollutants COD and TSS were chosen because UK wastewater management companies use COD and TSS concentrations in the wastewater, as part of the Mogden formula, when calculating wastewater disposal cost (Tools, 2019). The rate that wastewater companies charge for wastewater disposal differs between companies, see Fig. 1. This will have implications when determining whether it is more sustainable to maximise pollutant removal or energy generation. The energy generation potential of the H^{2}AD bioprocess is represented by the volume of methane produced. The H^{2}AD bioprocess produces biogas, made up of methane, carbon dioxide, oxygen and hydrogen sulphide. Of these, only methane has significant calorific value and can be used to estimate the energy potential from the biogas produced (Bauer et al., 2010).

This research aims to investigate the best techniques to develop a multiple target model of a process manufacturing system and demonstrate how that multiple target model may be utilised to increase sustainability within the system. The main contributions presented in this paper include:

- Evaluating the performance of two multiple target modelling techniques (1: problem transformation and 2: algorithm adaptation, detailed in Section 2) when modelling a process manufacturing system.
- Proposing a novel multiple target model by modifying algorithm adaptation models to use an artificial neural network (ANN) as a base model for the first time. The new model is then evaluated against existing multiple target modelling techniques.
- Utilising the final multiple target model to investigate the most sustainable manufacturing configuration of a novel industrial bioprocess for different manufacturing environments (e.g. feedstock characteristics and composition, regulations/legislation, waste management bodies).

2. Overview of multiple target modelling techniques

There are two main approaches for using DDMs as a base method for multiple target learning: *problem transformation* and *algorithm adaptation* (Borchani et al., 2015). Problem transformation is the transformation of a multiple target model into multiple single models, each solved independently in parallel using traditional DDM procedures (Borchani et al., 2015). Whereas, algorithm adaptation methods adapt the DDM algorithm to fit and/or

predict all outputs simultaneously in one DDM (Borchani et al., 2015). Algorithm adaption methods have been shown to perform better than problem transformation methods (Melki et al., 2017; Kocev and Ceci, 2015). This is accredited to their ability to capture not only the relationships between a model's inputs and outputs but also between the different outputs (Melki et al., 2017). A feature not possible in problem transformation models, because a single independent model is trained for each output (Borchani et al., 2015).

There are several approaches to use when building algorithm adaptation multiple target models, including linear target combinations for multiple target regression (Tsoumakas et al., 2014), multi-objective random forests (Kocev et al., 2007), multiple output artificial neural networks (Curteanu et al., 2011), boosted-neural network ensemble (Hadavandi et al., 2015) and multi-task deep learning (Ranjan et al., 2019). Recently explored was the possibility of stacking or chaining regression models together, with regression chains (RCs) proving the most effective of the two (Spyromitros-Xioufis et al., 2016). In building a RC model a random chain, or sequence, of the output variables is selected and for each output in the chain, single target models are built sequentially by using the output of the previous model as input for the next. The RC is an example of an algorithm adaptation multiple target modelling technique, because a RC uses selected model outputs to predict the additional model outputs (Fig. 2-A). Problem transformation methods create multiple independent models for each output and fail to capture the relationships between model outputs (Fig. 2-B) (Borchani et al., 2015).

Despite the previously discussed advantages of using an algorithm adaption model over a problem transformation model, there are two main problems with using a RC as an algorithm adaption model:

1. The relationships between the output variables at the start of the chain and the end of the chain are not exploited (Melki et al., 2017).
2. Prediction error at the start of the chain will propagate through the rest of the chain (Spyromitros-Xioufis et al., 2016).

To overcome these problems, an ensemble of regressor chains (ERC) was proposed (Spyromitros-Xioufis et al., 2016). The ECR works by creating multiple RCs for every possible permutation of the output sequence order. For example, the RC displayed in Fig. 2 (A) is ordered Y1−Y2−Y3, an ERC model would include multiple RC for orders Y1−Y3−Y2, Y2−Y1−Y3, Y2−Y3−Y1, Y3−Y1−Y2 and Y3−Y2−Y1. The final prediction values are obtained by taking the mean of the predicted values. Because of the computational time for building models with outputs that exceed 10 permutations, the sequence order is randomised, and 10 RCs are selected for constructing the ERC. As the number of output variables increases, the number of possible chains increases by a factorial function. Therefore, there is no guarantee that the 10 random chains generated will truly capture the relationships among the outputs. Additionally, building an ERC requires exponentially greater computational power as the number of outputs increases. Melki et al. proposed an adaption to the ERC, which instead used a single chain based on the maximisation of the correlations among the output variables (Melki et al., 2017). Melki et al. referred to their technique as a support vector regressor correlation chain, as it was built using a support vector regressor (SVR) algorithm as a base model. A SVR solves a non-linear problem by transforming the non-linearity between features and target using linear mapping (kernels) (Yildiz et al., 2017). They are highly effective algorithms, even when presented with small quantities of data (Yildiz et al., 2017). For this work, it is referred to as the maximum correlation chain
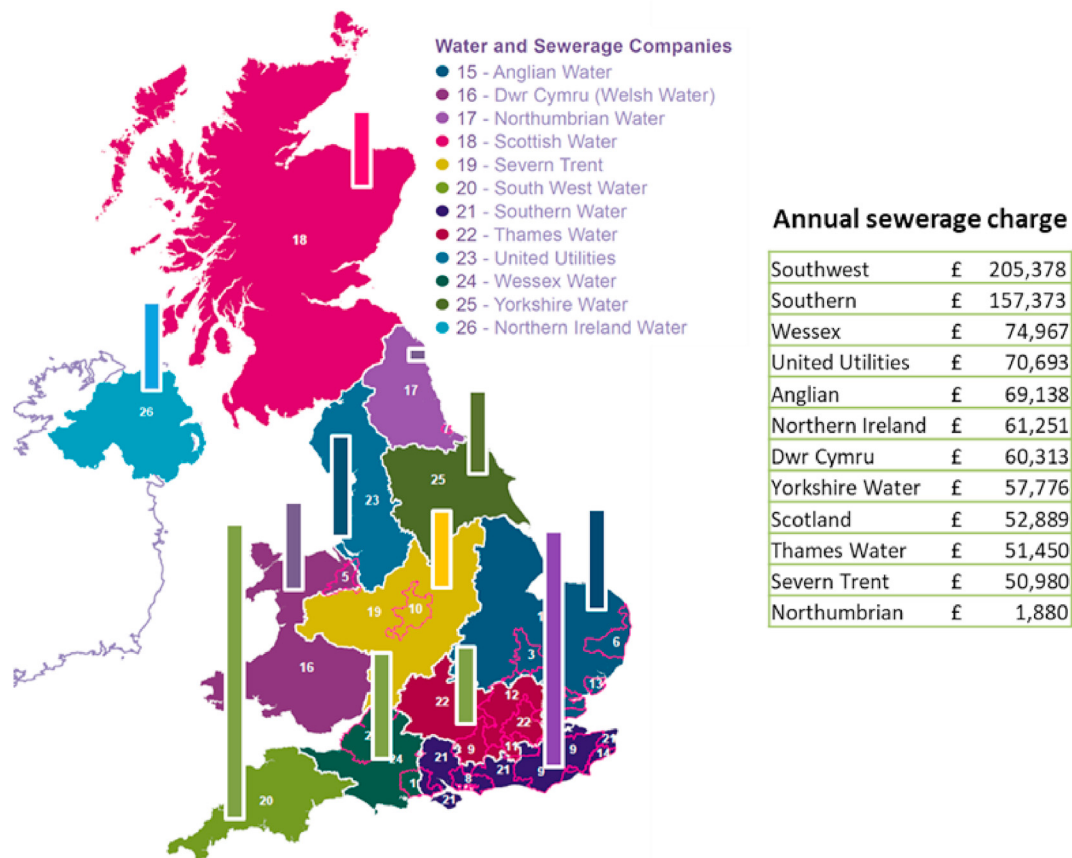
Fig. 1. Annual wastewater charge for different UK water and sewerage companies, based on 4 m³/d of wastewater with a chemical oxygen demand of 85,000 mg/L and total suspended solids of 984 mg/L using WRAP Mogden calculator (Tools, 2019).

(MCC) when applying the technique to other base models. To construct a MCC the first step is to calculate the outputs' correlation coefficient matrix, which will describe the linear relationship among the output variables. This can then be sorted into descending order, creating the MCC. The MCC then follows the same training procedure as a standard RC.

Problem transformation and algorithm adaptation multiple target models use DDM algorithms as a base model for each model developed in the stack or chain. Previous examples of ERC and MCC models have used random forest (RF) and SVR based models (Melki et al., 2017; Spyromitros-Xioufis et al., 2016). Random forests are an ensemble technique that uses multiple decision trees and a statistical technique called 'bagging' (Ahmad et al., 2017). Rather than just averaging predictions from multiple trees, an RF instead randomly samples from the training data for each tree and randomly subsets the input variables at splitting nodes until the minimum node size is reached (Ahmad et al., 2017). Artificial neural networks (ANNs) are a DDM algorithm that has yet to be utilised as a base model for ERC or MCC (Yildiz et al., 2017). Artificial neural networks are mathematical models that have been inspired by the connections made between neurons in a biological brain (Kim, 2017). The neural network is composed of nodes (which represent the neurons) and the connection between these nodes are given weight values to mimic how the brain alters the association between neurons (Kim, 2017). An ANN consists of an input layer, a single or multiple hidden layers, and an output layer. Artificial neural networks have excellent capabilities when modelling non-linear systems and are the most extensively utilised DDM in process manufacturing (Himmelblau, 2000).

The final multiple target modelling technique evaluated as part of this work is a multiple target artificial neural network (MT-ANN). An ANN model is highly suited to multiple target modelling as its architecture can be quickly adapted for multiple outputs (Spyromitros-Xioufis et al., 2016), as shown in Fig. 3. Because of this, MT-ANNs are among some of the earliest examples of multiple target modelling (Baxter, 1995; Caruana et al., 1995). An example is presented by Curteanu et al. who modelled the electrolysis process in wastewater treatment and found that an MT-ANN outperformed multiple individual ANN models (Curteanu et al., 2011). To the authors best knowledge, there has yet to be a comparison of the abilities of a MT-ANN with either the ERC or MCC modelling techniques and this article is the first to do so.

Recent work has shown the superiority of ERCs when compared to alternative multiple target methodologies, such as regressor stacking and deep structure stacking (Melki et al., 2017; Masmoudi et al., 2020; Barbon Junior et al., 2020). For example, Junior et al. developed an ECR model to predicting wheat flour quality parameters from near-infrared (NIR) spectroscopy data (Barbon Junior et al., 2020) and Masmoundi et al. developed an ECR model forecast multiple air pollutants simultaneously over two cities (Masmoudi et al., 2020). Both studies recommend the use of ECR as a tool to improve predictive performance. However, neither study demonstrates the application of multiple target models to investigate the relationship between the outputs or to evaluate different system configurations to optimise the one or more outputs. This is an omission this work aim resolve, by demonstrating a method by which process manufacturers can use multiple target models to improve their systems economic and environmental sustainability.
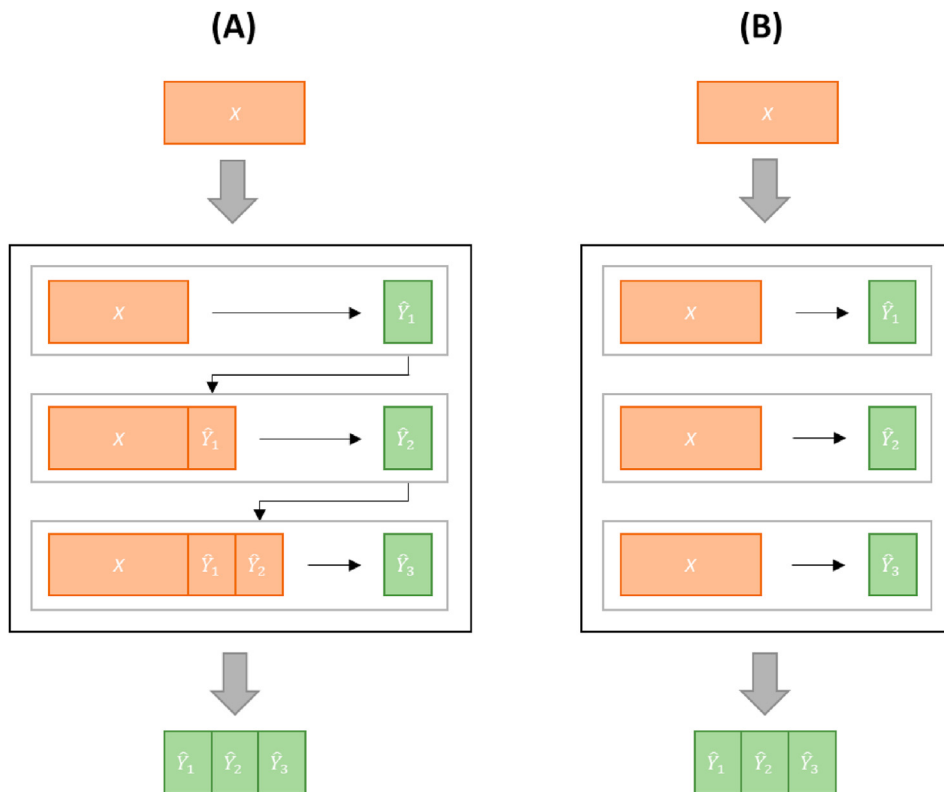
Fig. 2. Two examples of multiple target modelling: (A) algorithm adaptation and (B) problem transformation. Where the model input data (X) is represented by the orange box and the model output prediction ($\widehat{Y}_i$) by the green box.
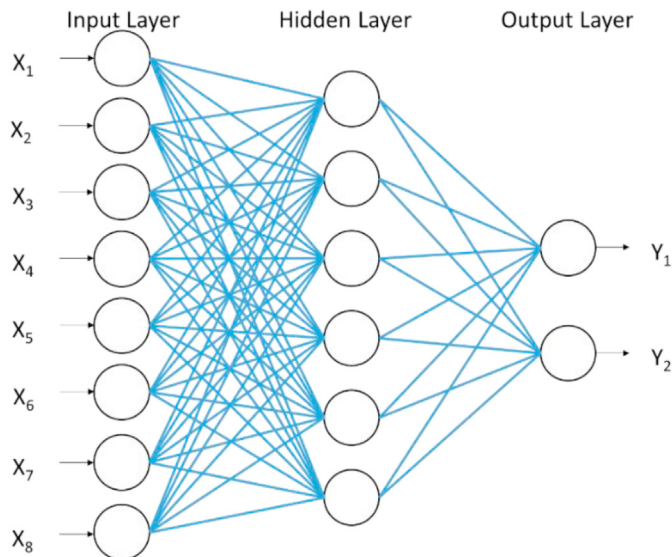

Fig. 3. Architecture of a multiple target artificial neural network (MT-ANN), with 8 inputs, 2 outputs and 1 hidden layer containing 6 neurons.

3. Material and methods

3.1. The H²AD process

The H²AD process used in this work is an industrial bioprocess plant developed by Lindhurst Engineering Ltd targeted at treating wastewaters from SMEs manufacturers spanning food and drink to industrial processing, due to its modular, low-cost design. Further detail about the H²AD technology is available in Section 1.1.

The H²AD plant treats wastewater generated by a dairy farm located in the East Midlands, UK. The farm waste contains cattle slurry, bedding waste, waste milk, footbath, parlour washings, and rainfall, which is separated by a screen press and the resulting liquid wastewater stored in a 3,000 m³ slurry tank. The wastewater from the slurry tank is fed into the H²AD system that generates bioenergy for the farm and improves the quality of that wastewater to support reuse (Fig. 4).

3.2. Model data

The model is developed from input and output data, illustrated in Fig. 5. The model input data originates from two sources, the H²AD feedstock data and the H²AD process data. To characterise the H²AD feedstock, a total of 18 water quality parameters were measured. In addition to this, the H²AD unit automatically-collected data on four H²AD process conditions via sensors wirelessly connected to a database hosted on a cloud platform. These were hydraulic retention time (HRT), pump speed, system temperature and system pressure. The H²AD outputs were characterised by three variables, %COD removal, %TSS removal and methane generation.

From previous work, there are 30 data-points available of water quality analysis performed on the farm waste samples, taken from the slurry tank feeding the H²AD system. In addition, there is one year's worth of historical H²AD process data automatically collected by the unit. This informed the decision to generate an additional synthetic 22 WQA data points bringing the total to 52 data points, representing one year's worth of weekly sampling. Synthetic data is used within modelling research to demonstrate and evaluate new modelling techniques (Bishop, 2006). Due to limited data available
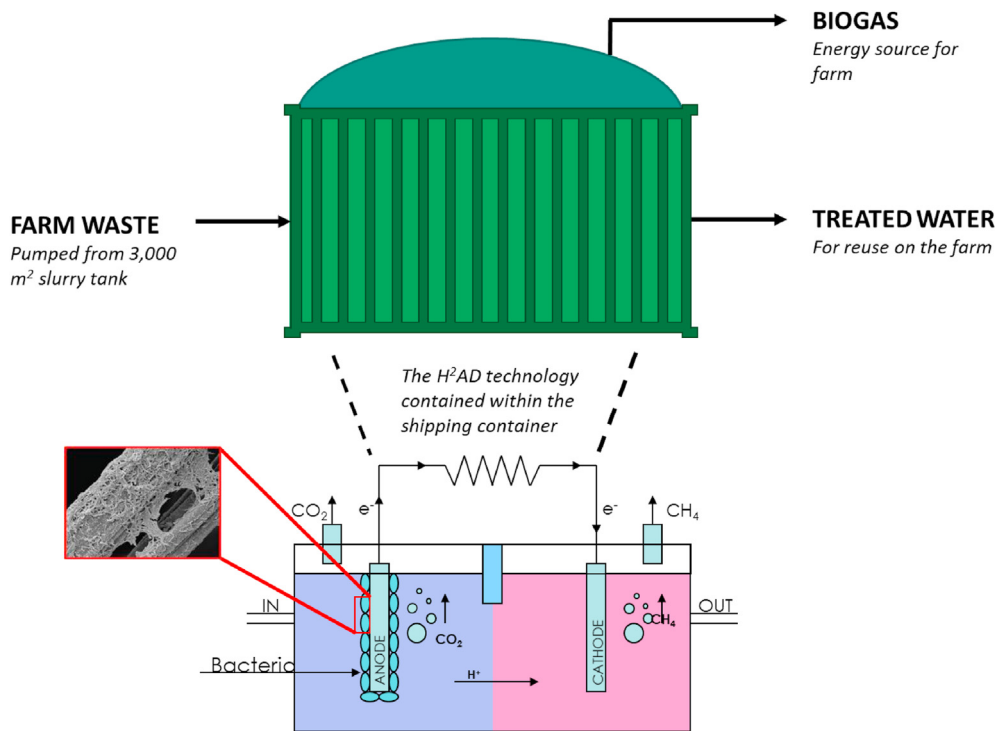
Fig. 4. The H^{2}AD manufacturing system (figured adapted from (H2AD, 2020)).
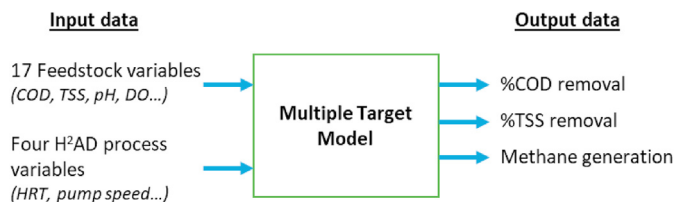

Fig. 5. The H2AD bioprocess model [chemical oxygen demand: COD; total suspended solids: TSS; dissolved oxygen: DO; hydraulic retention time: HRT; percentage of COD removal: %COD removal, percentage of TSS removal: %TSS removal, daily volume of methane generated: CH$_4$ generation].

from the H2AD bioprocess using the farm wastewater as a feed-stock, synthetic data was also developed to aid in demonstrating the application of multiple target models to increase the sustainability of process manufacturing systems. The farm wastewater data available providing information about the variability observed in the pollutant composition over time. While this is a validated methodology for evaluating new modelling techniques, it should be remembered that the conclusions drawn from the models are not always an accurate description of the H^{2}AD bioprocess.

When generating synthetic data the standard method is to randomly sample from a probability density function (PDF) fitted to existing data (Albuquerque et al., 2011). Pair-copula constructions (PCCs) are a popular tool for modelling multivariate data and creating synthetic data, due to their simple structure and high flexibility (Hobæk Haff et al., 2010). Comparisons of the mean values for the original and synthetic datasets are given in Table 1.

Due to insufficient data relating the model's input and output variables, it was not possible to follow the same PDF method. The approach utilised for generating synthetic output data followed the method of Brissette et al., whereby trends identified from historical H^{2}AD data governed the development of non-linear equations to generate synthetic output data from the input data (Brissette et al.,

2007). The features of the new dataset were compared against the original to assess how successfully the synthetic data captured the relationships in the original.

3.3. Data preparation, pre-processing and partitioning

For this study, the dataset was normalised to ensure variables of higher magnitudes are not given greater weight by the DDM. The COD and TSS data have magnitudes within the order of 10,000, while for DO the magnitude is much lower and in the order of 0.1. To normalise the data the minimax function was applied, as this normalises the variables without any loss of information (Al-Fattah et al., 2009). The equation to normalise the data is given below:

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

Where x is the variable, x_{norm} is the normalised variable and x_{min} and x_{max} is the minimum and maximum values of the variable being normalised respectively.

This work utilised a principal component analysis (PCA) to reduce the number of dimensions in the data to mitigate the risk of overfitting. Overfitting is the generation of a model that corresponds too closely or exactly to the noise (error) within the dataset, which negatively impacts future predictions (Srivastava et al., 2014). Principal component analysis is a mathematical procedure that transforms potentially correlated data into an orthogonal system of linearly uncorrelated principal components (PCs) (Abbas et al., 2018). All the PCs are orthogonal to each other and ordered so that the maximum variance is captured in PC1 and the variance decreases as one moves down the PCs. The standard practice is to reduce the number of variables so that 95% of the variance within the dataset is captured by the PCs (Abbas et al., 2018). This was the method followed in this work. By applying PCA to the H^{2}AD dataset, the input variables were reduced from 21 to 9 PCs.

Table 1
Summary of the model input data, including a comparison between the original farm wastewater quality dataset and the expanded synthetic dataset [dissolved oxygen: DO; conductivity: Cond.; chemical oxygen demand: COD; total suspended solids: TSS; total carbon: TC; total inorganic carbon: TIC; total organic carbon: TOC; total nitrogen: tN; nitrite: NO_2^-; nitrate: NO_3^-; chloride: Cl^-; phosphate: PO_4^{3-}; sulphate: SO_4^{2-}; zinc: Zn; hydraulic retention time: HRT; standard deviation: SD].

	Mean (SD) of original dataset	Mean (SD) of synthetic dataset	Range
pH	7.49 (0.134)	7.51 (0.185)	7.22−7.90
DO (mg/L)	0.187 (0.0376)	0.192 (0.0502)	0.110−2.80
Cond. (mS/mc)	11.8 (0.600)	11.7 (0.693)	10.3−12.8
TSS (mg/L)	17800 (5350)	19000 (5140)	11100−29200
COD (mg/L)	25400 (5590)	26600 (5500)	17400−38200
TC (mg/L)	8740 (2470)	9090 (2770)	4560−14100
TIC (mg/L)	1370 (367)	1440 (362)	828−2160
TOC (mg/L)	7350 (2160)	7650 (2500)	3540−12200
tN (mg/L)	1620 (477)	1800 (788)	770−3240
NH4+ (mg/L)	873 (178)	965 (267)	604−1490
NO2- (mg/L)	10.8 (2.22)	11.7 (2.44)	8.40−16.9
NO3- (mg/L)	77.0 (23.5)	84.7 (22.5)	51.0−131
Cl- (mg/L)	1540 (365)	1800 (494)	1190−3020
Water hardness (mg/L)	199 (35.6)	209 (35.1)	160−281
PO43- (mg/L)	295 (96.0)	326 (91.5)	179−529
SO4-2 (mg/L)	4890 (1310)	5330 (1350)	3480−8170
Zn (mg/L)	25.6 (7.17)	24.1 (7.90)	10.3−35.8
HRT (mg/L)	3.63 (0.877)	N/A	2.25−6.09
Pump speed (L/min)	9.68 (0.415)	N/A	8.84−9.94
System temp. (°C)	33.10 (1.99)	N/A	29.6−39.2
System pressure (bar)	0.00331 (0.00514)	N/A	0.00300−0.00500

The data used to develop the DDMs was then partitioned at several levels, as independent data is required to train, validate, test and evaluate the model. The data is first partitioned into model development data and unseen data. The model development data is further partitioned into training, validation and testing data (Bishop, 2006). Unseen data-points are kept apart until the models are completed to evaluate the prediction capabilities of the model on data not used during the model's development. To avoid sample representativeness issues, the semi-random partitioning framework developed by Liu et Cocea 2017 was employed to partition the unseen data points within and outside the model boundaries (Liu and Cocea, 2017). Ten points were taken from within the model development boundaries and six points outside the model boundaries.

The model development data was then portioned into training, validation and testing data. When developing a DDM, each algorithm has model parameters and hyperparameters that require tuning to find the optimal configuration. Hyperparameters are parameters that are set before being trained on a model, compared to model parameters that are learnt during training (Kim, 2017). For example, the number of neurons in an ANN's hidden layer is a hyperparameter; whereas, the weights between neurons is a model parameter. Validation data is used to tune the hyperparameters and training data is used to tune the model parameters. Testing data is utilised to evaluate the final model's ability to fit new data. The model development data points were randomised and 85% were partitioned for the training and validation of the model and the remaining 15% of data were partitioned for testing the final model. A Bayesian optimisation was performed to optimise the hyperparameters, using k-fold cross-validation to evaluate each configuration's ability to fit the validation data. For the k-fold cross-validation, a k value of 10 was chosen because of the limited (<100) number of data points (Charte et al., 2017). The model ability to fit new data was then evaluated using the testing data points. The model was also evaluated on unseen data to evaluate the model's capabilities to extrapolate beyond the model development data.

3.4. Multiple target models

With respect to modelling the capability of the H^2AD bioprocess

to reduce the pollutants (COD and TSS) from the wastewater and generate methane gas, 8 multiple target DDMs were built to fit the manufacturing data (Table 2). Three of the models were problem transformation models and five were algorithm adaption models. The DDM algorithms used RF, SVR and ANN as base models. Five algorithm adaption MT models were built to describe the H^2AD system. Two models were built following ERC techniques using SVR and ANN as base models, these are referred to as ERC-SVR and ERC-ANN respectively. The ECR models were built following the procedure outlined by Spyromitros-Xioufis et al. (2016). Two models were built following Melki et al. MCC procedure using SVR and ANN as base models, these are referred to as MCC-SVR and MCC-ANN respectively (Melki et al., 2017). The final MT model built was a MT-ANN. All the base models built were tuned by Bayesian optimisation and cross-validation procedures.

3.5. Model evaluation

The prediction capabilities of the various models were statistically evaluated in terms of the coefficient of determination (R^2), the normalised root of mean squared error (nRMSE) and the mean absolute percentage error (MAPE). The R^2 is the square of the sample correlation coefficient between observed values and predicted values, and is a measure of the explained variance of the model (Abbas et al., 2018). The nRMSE measures the normalised mean square magnitude of the error (Dalmau et al., 2015), while MAPE measures the absolute percentage of the errors (Hyndman and Koehler, 2006). The statistically best model is the one that has an R^2 closest to 1 while minimising nRMSE and MAPE. The R^2, nRMSE and MAPE are defined as follows:

$$R^2 = \frac{\left(\sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{\widehat{Y}}_i\right)\left(Y_i - \overline{Y}_i\right)\right)^2}{\sum_{i=1}^{n}\left(\widehat{Y}_i - \overline{\widehat{Y}}_i\right)^2\left(Y_i - \overline{Y}_i\right)^2} \tag{2}$$

Table 2
Multiple target models developed and their abbreviations.

Data-driven model	Abbreviation	Modelling approach
Random forest model	RF	Problem transformation
Support vector regressor	SVR	Problem transformation
Artificial neural network	ANN	Problem transformation
Ensemble regressor chain − support vector regressor	ERC-SVR	Algorithm adaption
Ensemble regressor chain − artificial neural network	ERC-ANN	Algorithm adaption
Maximum correlation chain - support vector regressor	MCC-SVR	Algorithm adaption
Maximum correlation chain − artificial neural network	MCC-ANN	Algorithm adaption
Multiple target artificial neural network	MT-ANN	Algorithm adaption

$$nRMSE = \frac{1}{Y_{max} - Y_{min}} \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2}{n}} \qquad (3)$$

$$MAPE = 100 \times \frac{\sum_{i-1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|}{n} \qquad (4)$$

Where n is the number of data points, $\widehat{Y}$ is the predicted value, Y_i is the actual observed value, and the symbol $_{-}$ is the average of the related values.

A ranking system was employed to determine the best overall model at predicting the H^{2}AD bioprocess outputs. The average statistical error metrics R^2, nRMSE and MAPE for each model were determined. Then for each average statistical error metrics, the models were ordered from best to worst and scored (top of the list scored one and the bottom scored three). The model's rank was then determined by summing together each model's scores for each error metrics and ordering the model's from lowest to highest.

4. Results and discussion

4.1. Modelling the H^{2}AD bioprocess

4.1.1. Problem transformation multiple target models
The capability of the base models (RF, SVR and ANN) to predict the %COD removal, %TSS removal and methane generation were statistically evaluated using R^2, nRMSE and MAPE (Table 3). The RF model is inferior at predicting all of the H^{2}AD outputs especially the %COD removal (R^2 of 0.63, nRMSE of 0.06 and MAPE of 8.56). The SVR and ANN show similar capabilities when predicting the %COD removal, though the ANN model has a slightly poorer nRMSE score (nRMSE of 0.05-ANN compared to 0.04-SVR). This indicates that the ANN model is subject to a greater number of outliers within the predictions. This trend between the SVR and ANN model is repeated when predicting methane generation (Table 3). However, when predicting %TSS removal the SVR model performs consistently worse than the ANN model (R^2 of 0.90, nRMSE of 0.08 and MAPE of 4.14, compared to R^2 0.96, nRMSE of 0.05 and MAPE of 3.22).

The results from averaging each model's R^2, nRMSE and MAPE and model ranking are displayed in Fig. 6. The average R2, nRMSe and MAPE scores indicate that the ANN model is superior at predicting the H^{2}AD bioprocess outputs compared to the RF and SVR model (Fig. 6).

Problem transformation models have been shown to be less accurate when predicting multiple outputs, as they are unable to capture the relationship between model outputs (Borchani et al., 2015). However, they can be effective at determining the suitability of DDM algorithms as a base model for algorithm adaption

Table 3
Comparison of the random forest (RF), support vector regression (SVR) and artificial neural network (ANN) models ability to fit the model development data [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; chemical oxygen demand: COD; total suspended solids: TSS; methane: CH$_4$].

Model	Output	R2	nRMSE	MAPE
RF	COD	0.63	0.06	8.56
	TSS	0.90	0.07	5.54
	CH4	0.95	0.07	2.53
SVR	COD	0.97	0.04	2.26
	TSS	0.90	0.08	4.14
	CH4	0.96	0.05	3.00
ANN	COD	0.96	0.05	1.80
	TSS	0.96	0.05	3.22
	CH4	0.96	0.06	2.36

models (Melki et al., 2017). After statistical analysis of the predictive capabilities of the RF, SVR and ANN models, the decision was taken to use only SVR and ANN algorithms as base models for the algorithm adaption multiple target models.

4.1.2. Algorithm adaption multiple target models
The ERC-SVR, ERC-ANN, MCC-SVR, MCC-ANN and MT-ANN models' capabilities to predict the %COD removal, %TSS removal and methane generation were also statistically evaluated using R^2, nRMSE and MAPE; the results are reported in Table 4. The results from ranking the models, displayed in Fig. 7, indicate that ERC-ANN is the superior model. The MCC-ANN and ERC-ANN have a similar MAPE score, indicating that on average they have similar prediction errors. However, the MCC-ANN has a higher average nRMSE error (0.046), which is a result of the model containing a greater number
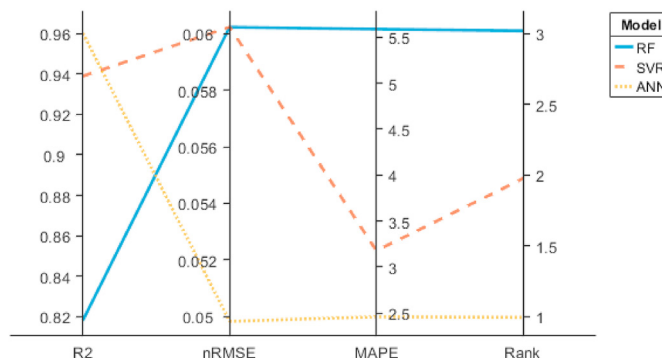
Fig. 6. Evaluation metrics of the problem transformation modes. Where the black line is the average performance of all three outputs, the blue line is percentage chemical oxygen demand removal output, the red line is percentage total suspended solids removal output and the yellow line is methane generation output. [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; random forest, RF; support vector regression, SVR; artificial neural network, ANN].

of outlier predictions. The ERC-ANN is an example of an ensemble model, which is able to mitigate the influence of outlier predictions. This is because the outputs from an ensemble model are an average of multiple predictions; an outlier may raise the value of the prediction, but by averaging the prediction an ensemble model is capable of reducing the outliers' impact on model performance (Levatić et al., 2017).
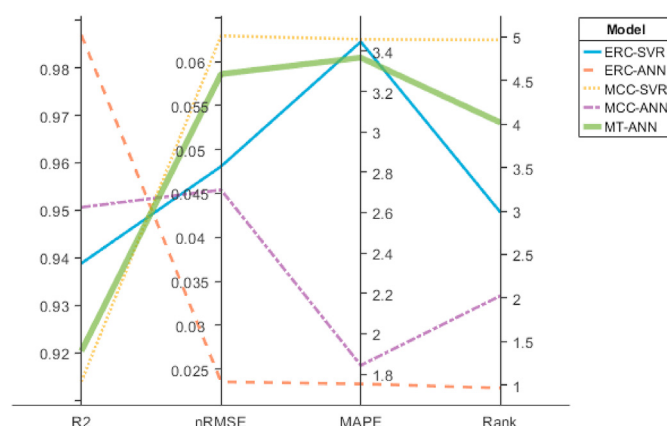
The RC models built using an ANN as a base model were ranked first and second, while the RC models using a SVR as a base model were ranked third and fifth. Similarly, the ERC modelling ranked higher than the MCC model when built with the same base model (ECR-ANN ranked first > MCC-ANN ranked second, ECR-SVR ranked third > MCC-SVR ranked fifth). This suggests that the ANN is a superior base model and that the ERC multiple target modelling technique was superior to the MCC technique. Of particular interest was comparing the ANN RC models to the MT-ANN model, as these techniques have never been compared before. The MT-ANN model was outperformed by both the ECR-ANN and MCC-ANN models, shown in Fig. 7. The reason for this may be that the models were constructed on only 36 data-points. Multiple target ANN models are known to be less capable when modelling systems where data is limited (Panerati et al., 2019). This is because, as an ANN architecture increases its requirement for data points to avoid overfitting (Al-Fattah et al., 2009). The RC ANN models built have smaller architectures (only one neuron in the outer layer), which may explain their better performance. Before drawing any conclusions on the multiple target models, it is important to evaluate the models' performance when making predictions for unseen data. Testing the models on unseen data was performed to evaluate a models' extrapolation capabilities beyond the data they were developed from which is important to determine the capability of the model for real-world application (Panerati et al., 2019).

Comparing the average statistical error scores of the SVR and ANN based single target models and the multiple target models show little difference in performance (averages for single target models: R^2 of 0.95, nRMSE of 0.06 and MAPE of 2.80; averages for multiple target models: R^2 of 0.94, nRMSE of 0.05 and MAPE of 2.78). These results are not consistent with previous work that has shown multiple target models to outperform single target models (Melki et al., 2017; Kocev and Ceci, 2015). To understand why this might have occurred, the analysis is broken down to compare the single and multiple target models for each algorithm. The ANN

Fig. 7. Evaluation metrics of algorithm adaption modes. Where the black line is the average performance of all three outputs, the blue line is %COD removal output, the red line is %TSS removal output and the yellow line is CH4 output. [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; support vector regression, SVR; artificial neural network, ANN; ensemble of regressor chains, ERC; maximum correlation chain, MCC].

based multiple target models outperformed the single target ANN model (single target ANN model: R^2 of 0.96, nRMSE of 0.05 and MAPE of 2.46; averages for ANN based multiple target models: R^2 of 0.97, nRMSE of 0.04 and MAPE of 1.79). These results agree with the previous work that capturing the relationship between outputs increases a model's performance (Melki et al., 2017; Kocev and Ceci, 2015). However, the same is not true for the SVR based multiple target models. The SVR based multiple target models were outperformed by the single target SVR model (single target ANN model: R^2 of 0.94, nRMSE of 0.06 and MAPE of 3.13; averages for SVR based multiple target models: R^2 of 0.93, nRMSE of 0.06 and MAPE of 3.46). This may be due to the SVR reduced capability to predict the %TSS removal (R^2 of 0.90) compared to %COD removal (R^2 of 0.97) and methane generation (R^2 of 0.96). It appears that for the SVR-based multiple target model, the benefit of capturing the relationship between outputs is outweighed by %TSS removal prediction error that is propagated throughout the model.

4.1.3. Comparison of the models' performance on unseen data

A common error when developing a DDM, is to only optimise the model's capability to fit the data used in the model development (Panerati et al., 2019). The best DDMs are able to extrapolate beyond the data they were developed from, to make accurate predictions on unseen data. This is important if a process manufacturer wants to utilise the model for predictions on future data, forecasting how the system may react to change, and to assess how much of the process manufacturing system's variability is captured in the model. The 16 unseen data points were partitioned from the manufacturing data to represent the parameter space both inside and outside the system. The prediction capabilities of the multiple target models for unseen data were statistically measured in terms of R^2, nRMSE and MAPE. Table 5 reports the statistical comparison of all models for each of the H^2AD bioprocess's outputs.

When evaluating the models' abilities to make predictions on unseen data within the system, it is of interest to note that the ECR is no longer clearly superior to the MCC technique; for example, both ERC-SVR and MCC-SVR have a R^2 of 0.75 for inside TSS predictions. The ERC-ANN model is no longer a superior model, due to its relatively poor performance when predicting the %COD removal (R^2 of 0.33, nRMSE of 0.17 and MAPE of 10.17). The ERC-SVR, ERC-ANN and MCC-ANN average performance are all relatively equal when making predictions for unseen data inside the system, each

Table 4
Comparison of the ensemble of regressor chains (ERC) and maximum correlation chain (MCC) models using support vector regression (SVR) and artificial neural network (ANN) based models with a multiple target ANN (MT-ANN) ability to fit the model development data. [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; chemical oxygen demand: COD; total suspended solids: TSS; methane: CH_4].

Model	Output	R2	nRMSE	MAPE
ERC-SVR	COD	0.92	0.04	3.92
	TSS	0.92	0.07	3.97
	CH4	0.97	0.03	2.46
ERC-ANN	COD	0.98	0.02	1.53
	TSS	0.99	0.03	2.20
	CH4	0.99	0.02	1.50
MCC-SVR	COD	0.91	0.05	3.13
	TSS	0.91	0.08	4.43
	CH4	0.92	0.06	2.85
MCC-ANN	COD	0.94	0.05	1.83
	TSS	0.96	0.05	1.79
	CH4	0.95	0.04	1.89
MT-ANN	COD	0.92	0.05	3.03
	TSS	0.91	0.08	3.93
	CH4	0.94	0.05	3.19

Table 5
Comparison on unseen data of the ensemble of regressor chains (ERC) and maximum correlation chain (MCC) models using support vector regression (SVR) and artificial neural network (ANN) based models with a multiple target ANN (MT-ANN) [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; chemical oxygen demand: COD; total suspended solids: TSS; methane: CH_4].

| Model	Output	R2		nRMSE		MAPE	
		Inside	Outside	Inside	Outside	Inside	Outside
ERC-SVR	COD	0.47	0.50	0.06	0.11	9.21	22.75
	TSS	0.75	0.88	0.12	0.11	6.26	12.72
	CH4	0.63	0.15	0.10	0.17	7.34	18.27
ERC-ANN	COD	0.33	0.82	0.17	0.10	10.17	16.23
	TSS	0.80	0.91	0.09	0.09	5.02	10.78
	CH4	0.75	0.36	0.08	0.14	6.58	16.10
MCC-SVR	COD	0.35	0.55	0.08	0.12	8.72	22.31
	TSS	0.75	0.90	0.12	0.11	7.34	11.71
	CH4	0.43	0.11	0.13	0.19	11.53	18.28
MCC-ANN	COD	0.62	0.55	0.13	0.16	8.23	21.40
	TSS	0.63	0.93	0.11	0.11	7.65	7.02
	CH4	0.82	0.49	0.06	0.14	7.25	14.43
MT-ANN	COD	0.00	0.85	0.16	0.14	16.01	11.51
	TSS	0.40	0.50	0.21	0.33	10.89	17.69
	CH4	0.46	0.41	0.13	0.18	7.97	17.53

Fig. 8. Statistical analysis of algorithm adaption MT modes performance on unseen data. Where the black line is the average performance of all three outputs, the blue line is %COD removal output, the red line is %TSS removal output and the yellow line is CH4 output. [Coefficient of determination, R^2; normalised root of mean squared error, nRMSE; mean absolute percentage error, MAPE; support vector regression, SVR; artificial neural network, ANN; ensemble of regressor chains, ECR; maximum correlation chain, MCC].

performing slightly better for a different statistical metric (ERC-SVR, nRMSE of 0.9; ERC-ANN, MAPE of 7.5; MCC-ANN, R^2 of 0.64). Although, the MT-ANN model remains significantly worse than all RC models (average R^2 of 0.29), emphasising the point raised earlier about the tendencies or larger ANN architectures to overfit. This highlights the advantages of the algorithm adaptions techniques developed by Spyromitros-Xioufis et al. and Melki et al. when building a multiple target model (Melki et al., 2017; Spyromitros-Xioufis et al., 2016).

The ranking system described in Section 3.2.1 was again used to systematically determine which model was superior when making predictions on unseen data both inside and outside of the system boundaries. The results in Fig. 8 show that the ECR-ANN model was the superior model. As the ECR-ANN has proved to be the statistically best model both when making predictions for the model development data and new unseen data, it was determined to be the best model to describe the three outputs of the H^2AD bioprocess. The ECR-ANN model was then used to increase the sustainability of the H^2AD bioprocess.

4.2. Utilising the final data-driven model to optimise H^2AD bioprocess sustainability

The H^2AD bioprocess can be manipulated to promote improved performance for one output over the other by changing the set point of the H^2AD process conditions (HRT, pump speed, system temperature and system pressure). To better understand the effect of the H^2AD process conditions on the bioprocess, the ECR-ANN model was employed to generate three-dimensional surface plots. Fig. 9 and Figs. A1-5, show the effect of the four investigated H^2AD process conditions on the H^2AD bioprocess outputs (%COD removal, %TSS removal and methane generation). The plots were produced by using the ECR-ANN to predict the effect of two H^2AD process conditions on the H^2AD bioprocess outputs, while holding water quality parameters and the remaining two process conditions constant at intermediate values. The knowledge gained from the surface plots can inform the manufacturer how to optimise the bioprocess, dependant on the manufacturing environment's requirements, to find the most sustainable solution.
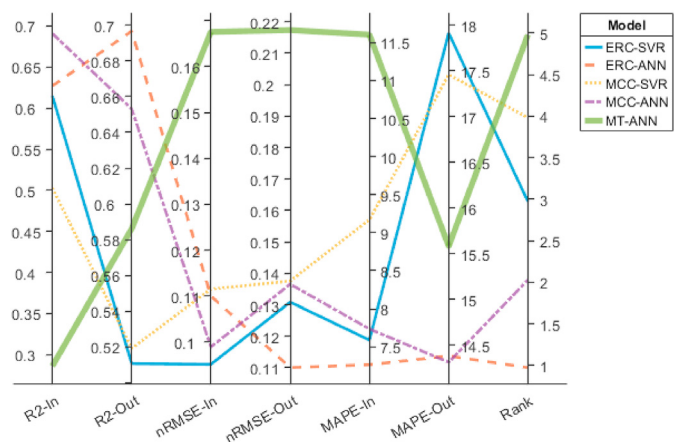
4.2.1. Understanding the relationship between process conditions and pollutant remediation and energy generation

The effect of HRT, pump speed, system temperature and system pressure on %COD removal by the H^2AD bioprocess are shown in Fig. 9 and Fig. A1. The contour plot, in Fig. 9 (A), shows that at lower pump speeds the removal of COD is driven by the HRT. However, when the pump speed is increased beyond 9.2 L/min the effect of pump speed on %COD removal increases. The contour plots in Fig. A1 (A) and (B) also show that at pump speeds of less than 9.2 L/min the pump speed has less effect on the %COD removal than the system temperature and system pressure respectively. If this was real data, rather than synthetic, it may be possible to hypothesise justifications for this observed phenomenon. For example, that pump speeds above 9.2 L/min cross a critical recirculation rate, within the reactor, that enables sufficient mixing of the fluid to reduce dead space within the reactor, increasing the reactor's efficiency. The effect of recirculation rates on the volume of dead space within reactors is a well-studied field (Fogler, 2017). The advantage of the ECR-ANN based data analysis is that it is able to identify the critical recirculation rate and is able to determine the benefits of increasing recirculation rate in comparison to the other H^2AD process conditions. This can help to inform manufacturers which process conditions have a more significant effect on a system.

Within the ranges investigated, HRT is the driving factor on %COD removal compared to both pump speed and system pressure. Therefore, HRT rates should be prioritised if aiming to control the %COD removal (Fig. 9 (A) and (C)). Whilst Fig. 9 (B) indicates that HRT and system temperature have a similar effect and that both should be given consideration when attempting to manipulate %COD removal. Generally, ECR-ANN based data analyses indicate that shorter HRT, higher pump speeds, system temperatures and lower system pressure increased the %COD removal by the H^2AD bioprocess from the wastewater feedstock. By repeating this analysis for other H^2AD outputs (%TSS removal and methane generation) it was possible to identify the value of each process condition that will maximise each output.

Table 6 reports the values of each process condition that will maximise each output. By comparing these values it becomes apparent that the outputs %COD removal and methane generation
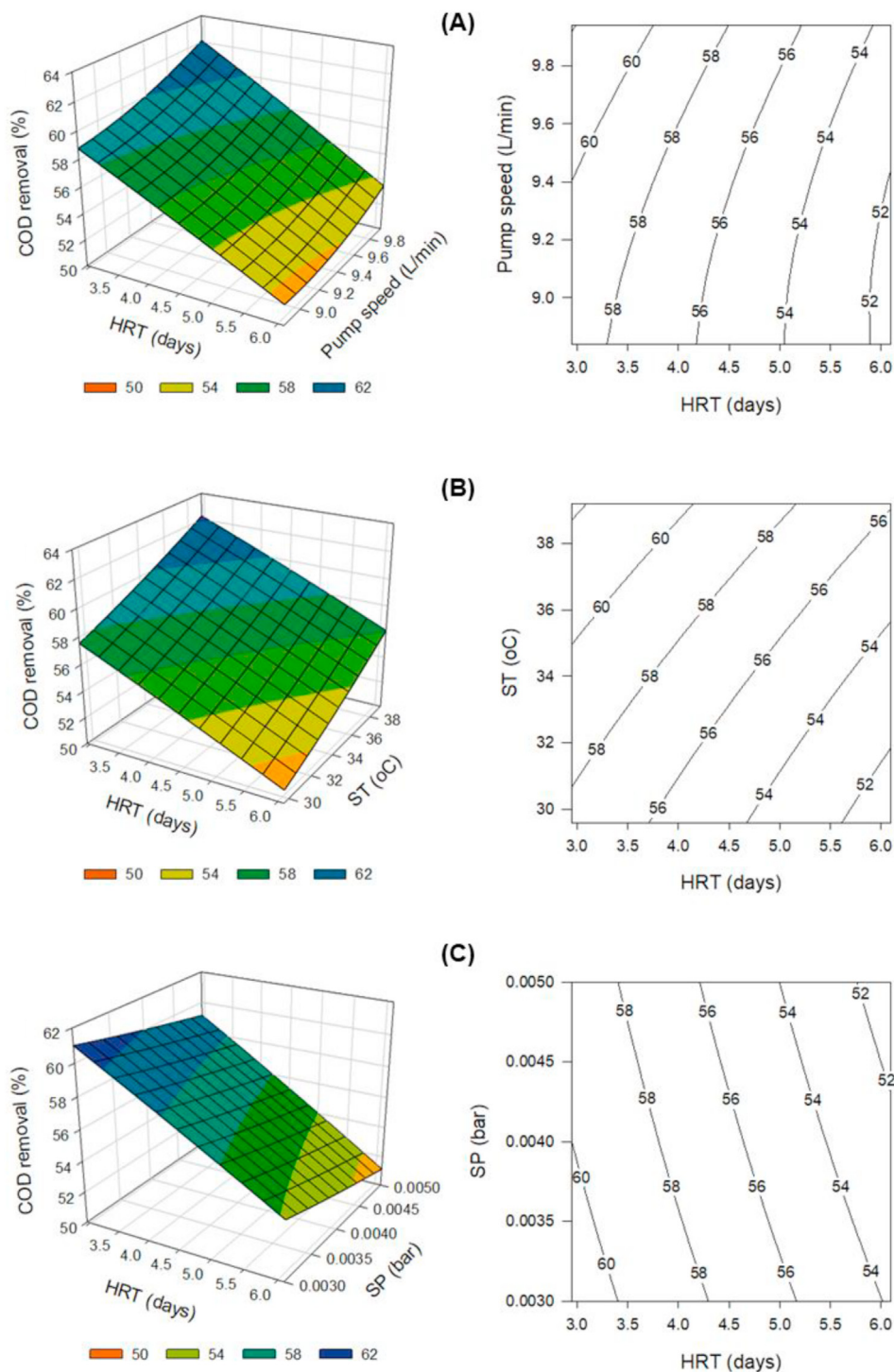
Fig. 9. Surface plots (left) and corresponding contour plots (right) showing the effects of the H^{2}AD process conditions on the chemical oxygen demand (COD) removal as predicted by the ECR-ANN model with set temperature (ST) and system pressure (SP) held constant (A), with pump speed and SP held constant (B) and with pump speed and ST held constant (C). [Hydraulic retention time, HRT].

have a similar configuration of process conditions to maximise each output. The only difference is that a high system temperature (39.2 °C) increases the percentage of COD removed, but decreases the volume of methane generated. Fig. 9 (B) and Fig. A4 (B) indicates that system temperature has a stronger effect on %COD removal than it does on methane generation. Therefore, a manufacturer aiming to maximise both should consider setting a higher

system temperature as this will have less of a negative effect on methane generation than a low system temperature will have on % COD removal. Manufacturers aiming to maximise all three H^{2}AD outputs are likely to face difficulty arising from conflicting optimal configuration of the HRT set point between %TSS removal and the other two outputs. The ECR-ANN based data analyses indicate that HRT is a driving force for controlling all three of the H^{2}AD outputs.

Table 6
Optimal process conditions values for each of the H^{2}AD bioprocess outputs. [Hydraulic retention time, HRT; set temperature, ST; system pressure, SP; percentage chemical oxygen demand removal, %COD removal; percentage total suspend solids, %TSS removal; volume of methane generated, CH$_4$ generation].

Process condition	%COD removal	%TSS removal	CH$_4$ generation
HRT (days)	2.95	6.10	2.95
Pump speed (l/min)	9.94	9.94	9.94
System temperature (°C)	39.20	39.20	29.6
System pressure (bar)	0.003	0.005	0.003

The %TSS removal output favours a longer HRT, while the other two outputs favour a shorter HRT. Selecting which HRT to operate will depend on the manufacturer's goal, the incoming wastewater COD and TSS composition and the manufacturing environment operating within, as illustrated in Section 3.4.2.

4.2.2. Determining the most sustainable configuration of the H^{2}AD bioprocess

After analysing the impacts of the process conditions on the H^{2}AD bioprocess, it became apparent that when optimising for % COD removal and methane generation, %TSS removal would decrease. Finding the 'best' solution will depend on maximising the economic and environmental benefits when treating a manufacturer's wastewater stream. The environmental objective of the system is to minimise the pollutants leaving the system (with removal from the waste/water feedstock via degradation/sorption) and maximise the energy in the form of biogas recovered from the waste/water. This closely matches the economic objective of the H^{2}AD system to (A) reduce waste disposal costs associated with manufacturing process by reducing the pollutant load in wastewater and (B) reduce the cost of energy by recovering energy from the waste stream. However, there are also economic costs associated with operating the H^{2}AD. These are largely from the required energy to heat the incoming wastewater to the required system temperature and maintaining system temperature. These may come into conflict with each other, as operating the H^{2}AD system on higher temperatures was found to increase both pollutant removal and energy generation. The extra cost associated with operating at higher temperatures may be greater than the economic benefits from the savings generated from the reduced pollutant load and onsite energy generation.

The most sustainable solution will differ between manufacturing environments (e.g. feedstock characteristics and composition, regulations/legislation, waste management bodies) and the feedstock (wastewater) characteristics. When the H^{2}AD effluent is being emitted to a waste management body that charges for pollutant load and is the driving economic force of the system, the goal would be to maximise pollutant removal regardless of bioenergy generation. As the results from Section 3.4.1 indicate that %COD removal and %TSS removal have different optimal conditions, the configuration of process conditions would depend on the characteristics of the wastewater being treated by the H^{2}AD system. However, if pollutant load charges in the effluent are lower the optimal economic solution may be to maximise energy recovery from the waste stream.

The ECR-ANN model was used to determine the economic benefit of optimising the H^{2}AD configuration dependant on wastewater composition and further treatment costs. Four different wastewater composition scenarios were proposed using different maximum and minimum values of COD and TSS concentrations observed in the model data, the values are presented in Table 7. The remaining wastewater feedstock characteristics were held constant at an intermediate value. The two H^{2}AD configurations were (1)

optimising for %COD removal and methane generation (configuration 1) and (2) optimising for %TSS removal. The H^{2}AD process conditions were taken for each configuration from the results of the surface plot modelling in Table 6.

The ECR-ANN model predicted the COD and TSS removal rates and the volume of biogas generated for each configuration for the four scenarios. The MAPE error metric was used to estimate error boundaries for the model's predictions. The removal rates were then used to calculate the pollutant concentration in the treated wastewater, using Equation (5). The maximum and minimum pollutant concentrations were calculated using the MAPE metric in Equation (6). The results for each scenario and H2AD configuration are presented in Table 7.

$$x_p = x - x \times r_p \tag{5}$$

$$x_p = x - x \times r_p \times (1 \pm MAPE) \tag{6}$$

Where x_p is the predicted pollutant concentration in the treated wastewater, x is the pollutant concentration in the untreated wastewater, r_p is the predicted pollutant removal rate by the H2AD bioprocess and MAPE is the mean absolute percentage error associated with the prediction.

was conducted to identify which H^{2}AD configuration was the most sustainable for each scenario depending on whether the wastewater was being sent to a Water Company that charge a high amount (United Utilities) or a low amount (Wessex Water). The disposal charges were then calculated for each scenario using the Mogden formula, assuming 4m^3/day of wastewater was being treated (Tools, 2019). The energy value from the methane gas was calculated by using equation (5) from Bauer et al. (2010) and then converted to kilowatt hours (KWH). A biogas generator is not capable of converting all this energy to electricity, as approximately 65% of the energy is lost as heat and other mechanical losses (Bauer et al., 2010). Therefore, the predicted energy generated by the H^{2}AD was reduced by 65% to account for inefficiencies. The average UK price of £0.124 per KWH at time of writing was used to determine the economic benefit from the biogas (UK Power, 2019).

$$1 \ m^3 \ of \ CH_4 = 39.79 \ MJ \tag{7}$$

The economic analysis of the ECR-ANN model results are reported in Table 8. The analysis indicates that for manufacturing environments that have a low disposal charge for H^{2}AD effluent, the optimal H^{2}AD configuration is the removal of COD and generation of CH$_4$. This is true for all the wastewater composition scenarios, though the largest economic benefit (£30.00 per day) from optimising the H^{2}AD configuration was for low COD and high TSS concentrations in the wastewater. Optimising the H^{2}AD configuration would save the process manufacturer £10,950 per annum, which reduces the wastewater treatment costs by 17.0%. When operating in a manufacturing environment with a high disposal charge, the H^{2}AD configuration also favoured COD removal and methane generation. There was one exception which was when the wastewater contained low COD and high TSS concentrations. In this instance by configuring the H^{2}AD to remove TSS would result in a £5.00 per day or £1,825 per annum economic benefit. Overall, this manufacturing environment saw the biggest economic benefit when the H^{2}AD was correctly configured to treat for high COD and high TSS concentration wastewater. By using a H^{2}AD configuration to optimise COD removal would benefit the process manufacturer by £5,475 per annum, which reduces the wastewater treatment costs by 4.5%.

Table 7
Chemical oxygen demand (COD) and total dissolved solids (TSS) concentration in the H^{2}AD input and output streams for each wastewater composition scenario for both H^{2}AD setups.

Scenario	Economic analysi		Configuration 1		Configuration 2	
	Pollutant	Pollutant concentration (mg/L)	Pollutant concentration (mg/L)	Methane generation (L/day)	Pollutant concentration (mg/L)	Methane generation (L/day)
1	COD	40,000	11900 (+/-19.5%)	5270 (+/-7.3%)	20500 (+/-7.9%)	4040 (+/-7.3%)
	TSS	30,000	24100 (+/-1.9%)		21000 (+/-3.3%)	
2	COD	40,000	12300 (+/-18.6%)	5250 (+/-7.3%)	20500 (+/-8.2%)	4130 (+/-7.3%)
	TSS	13,000	10600 (+/-1.8%)		9060 (+/-3.3%)	
3	COD	17,000	5070 (+/-19.3%)	5240 (+/-7.3%)	8740 (+/-7.8%)	4010 (+/-7.3%)
	TSS	30,000	24100 (+/-1.9%)		21000 (+/-3.3%)	
4	COD	17,000	5070 (+/-19.4%)	5120 (+/-7.3%)	8740 (+/-8.1%)	4000 (+/-7.3%)
	TSS	13,000	10600 (+1.8%, −1.7%)		9060 (+/-3.3%)	

5. Conclusion

When searching for the most sustainable manufacturing solution, multiple economic and environmental objectives must be considered. The objectives often come into conflict with one another. By developing a multiple target model of the manufacturing system, one can investigate the impact of various manufacturing factors on the objectives and understand how the objectives relate to one another. This will enable manufacturers to determine the most sustainable configuration of their process manufacturing systems. Despite recent advances in multiple target modelling, there are limited examples of multiple target models being developed for process manufacturing systems. The multiple target models that have been developed frequently use a problem transformation approach, where multiple independent single target models are developed for each output. This fails to capture the relationship between the model's outputs, resulting in models with inferior predictive capabilities.

In this study, two state-of-the-art multiple target modelling approaches were applied to a manufacturing system for the first time. These were regressor chain (RC) based techniques called ensemble of regressor chains (ERC) and maximum correlation chains (MCC). The RC technique treats the outputs of the system as inputs for subsequent outputs, thus capturing the relationship between model outputs. Previous work has used only random forest (RF) and support vector regressors (SVR) algorithms as base models for RC models. This study was the first example of an artificial neural network (ANN) based model being used for either an ERC or a MCC multiple target model. Previous work has shown the versatility of ANNs when modelling process manufacturing systems. This was reflected in this study, as the ANN was able to consistently outperform the ERC and MCC built from SVR base models. After statistical analysis of the model predictions on unseen data and after comparing the models' predictions distributions with that of the real data, the ECR-ANN model was found to be the superior model.

Utilising the ECR-ANN model to develop surface plots of %COD removal, %TSS removal and methane generation, it was found that the %COD and methane generation favoured closely similar configurations of the H^{2}AD system's process conditions. However, the %TSS removal was found to have the opposite relationship with all process conditions apart from the system temperature. This demonstrated the complex task when identifying the most sustainable solution, balancing both economic and environmental considerations. What is sustainable for one manufacturing environment may not be for another, as different environments have different economic and environmental objectives. Predictions from the ECR-ANN model were able to achieve economic benefits up of

Table 8
Economic analysis of the wastewater disposal costs before and post H2AD treatment and the economic benefit from the methane generated.

Wastewater company	Scenario	Disposal charge with no treatment	Disposal charge post H2AD	Saving	Value from methane	Total economic benefit
		£/day	£/day	£/day	£/day	£/day
Configuration 1						
United Utilities	1	415	240 (+/-5.8%)	175 (+/-8.1%)	10 (+/-7.3%)	**185** (+/-8.1%)
	2	285	140 (+/-8.6%)	145 (+/-8.2%)	10 (+/-7.3%)	**155** (+/-8.1%)
	3	310	210 (+/-3.8%)	100 (+/-8.0%)	10 (+/-7.3%)	110 (+/-7.9%)
	4	180	105 (+/-5.6%)	75 (+8.1%, −8.0%)	10 (+/-7.3%)	**85** (+8.0%, −7.9%)
Wessex Water	1	160	85 (+/-7.0%)	75 (+/-8.1%)	10 (+/-7.3%)	**85** (+8.1%, −8.0%)
	2	120	55 (+/-9.8%)	65 (+/-8.2%)	10 (+/-7.3%)	**75** (+/-8.1%)
	3	110	70 (+/-4.5%)	40 (+/-8.0%)	10 (+/-7.3%)	**50** (+/-7.9%)
	4	70	40 (+/-6.5%)	30 (+8.1%, −8.0%)	10 (+/-7.3%)	**40** (+8.0%, −7.9%)
Configuration 2						
United Utilities	1	415	255 (+/-4.9%)	160 (+/-8.0%)	5 (+/-7.3%)	165 (+/-8.0%)
	2	285	165 (+/-6.0%)	120 (+/-8.1%)	5 (+/-7.3%)	125 (+8.0%, −8.1%)
	3	310	205 (+/-4.2%)	110 (+/-7.9%)	5 (+/-7.3%)	**115** (+/-7.8%)
	4	180	110 (+/-5.0%)	70 (+/-8.0%)	5 (+/-7.3%)	75 (+/-7.9%)
Wessex Water	1	160	95 (+5.3%, −5.6%)	65 (+8.6%, −8.0%)	5 (+/-7.3%)	70 (+8.5%, −8.0%)
	2	120	65 (+/-6.3%)	50 (+/-8.1%)	5 (+/-7.3%)	55 (+/-8.0%)
	3	110	70 (+/-4.3%)	40 (+/-7.9%)	5 (+/-7.3%)	45 (+/-7.8%)
	4	70	45 (+5.2%, −5.1%)	25 (+8.0%, −8.1%)	5 (+/-7.3%)	30 (+/-7.9%)

£10,950 per annum.

CRediT authorship contribution statement

Oliver J. Fisher: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Nicholas J. Watson:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision. **Laura Porcu:** Investigation, Validation, Resources. **Darren Bacon:** Validation, Resources. **Martin Rigley:** Validation, Resources. **Rachel L. Gomes:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
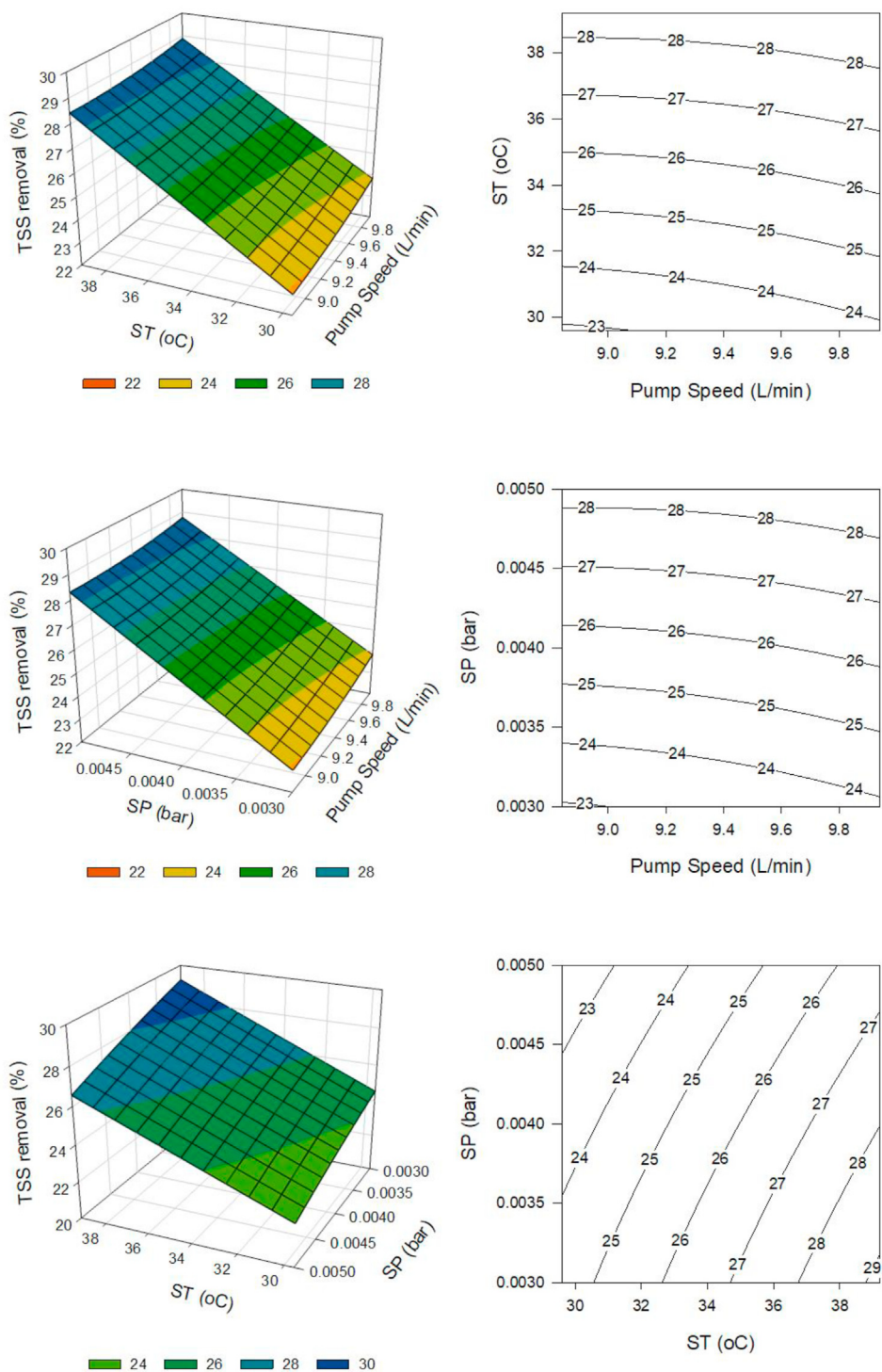
Appendix. (A): Surface plots
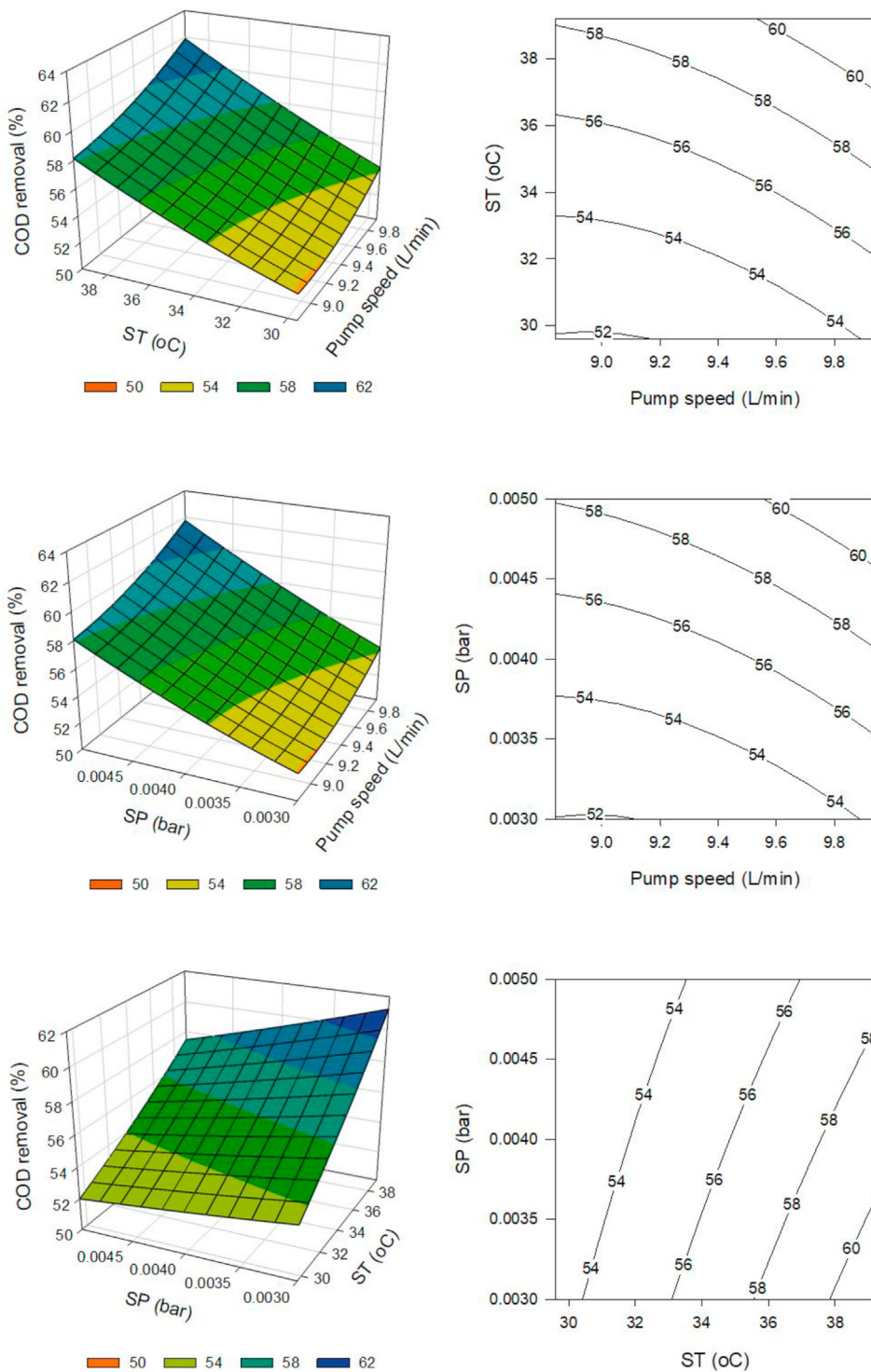
Fig. A.1. Surface plots (left) and corresponding contour plots (right) showing the effects of the H²AD process conditions on the chemical oxygen demand (COD) removal as predicted by the ECR-ANN model with hydraulic retention time (HRT) and system pressure (SP) held constant (A), with HRT and set temperature (ST) held constant (B) and with HRT and pump speed (C).
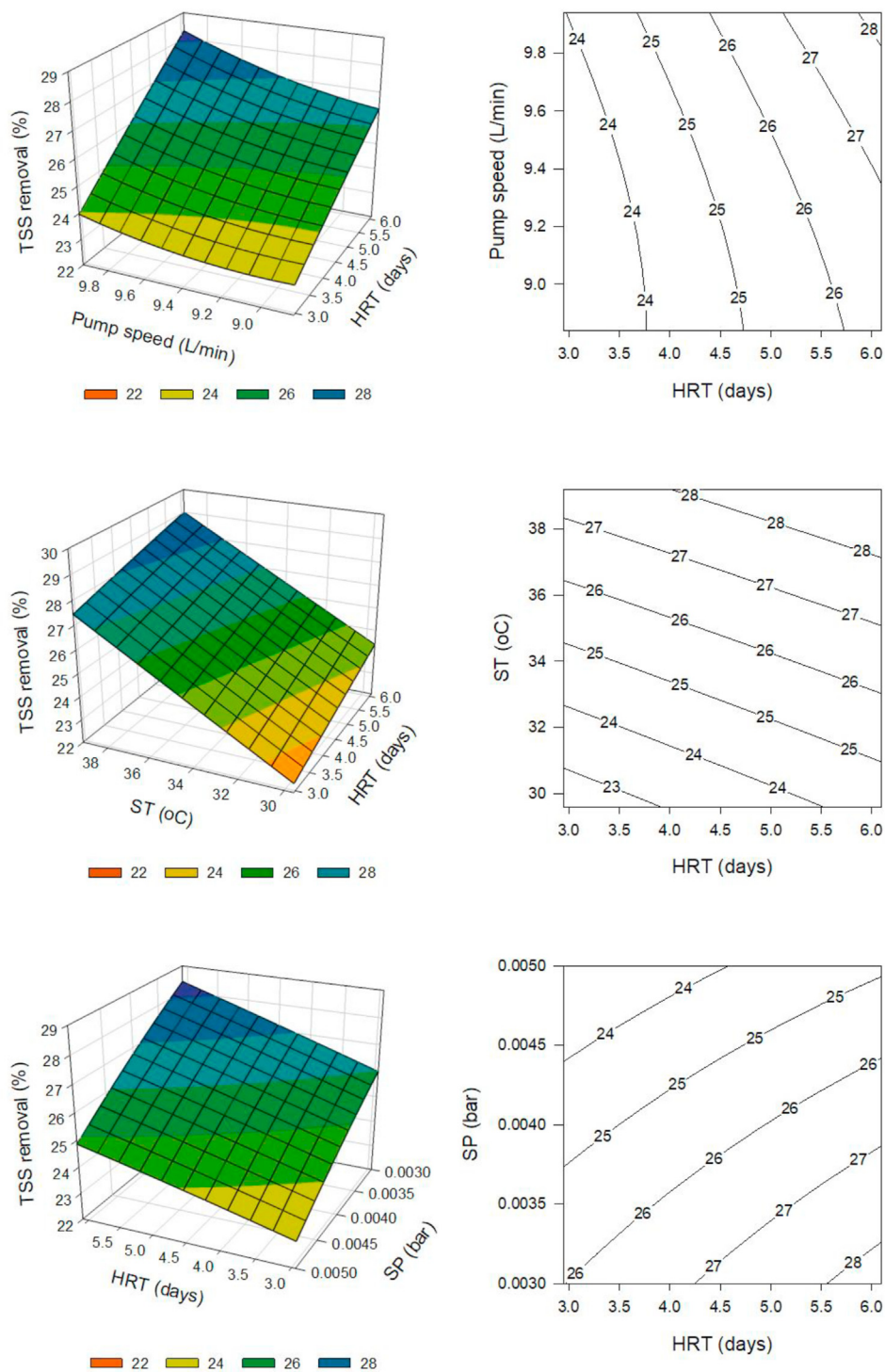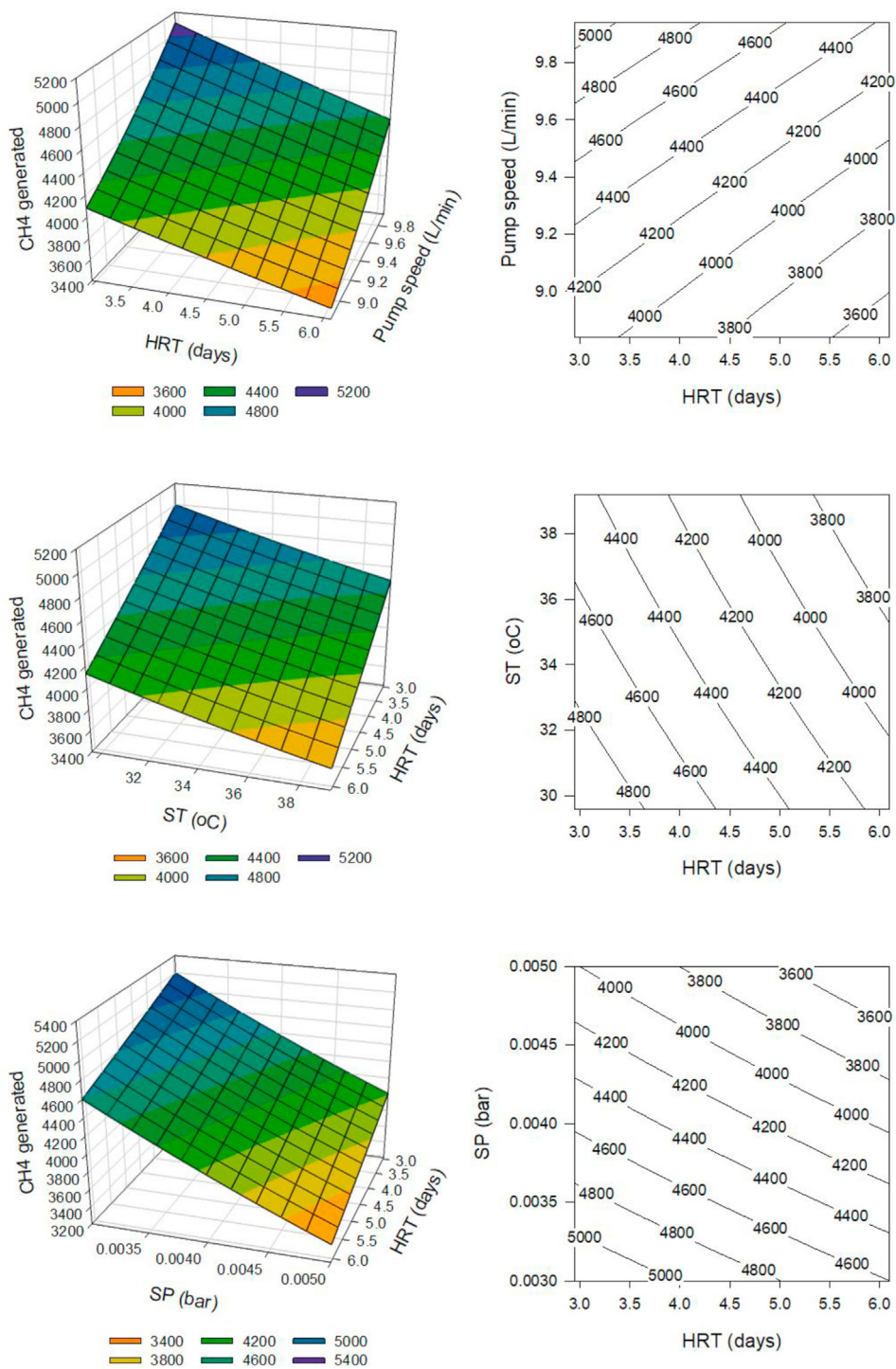
Fig. A.2. Surface plots (left) and corresponding contour plots (right) showing the effects of the H^{2}AD process conditions on the total suspended solids (TSS) removal as predicted by the ECR-ANN model with set temperature (ST) and system pressure (SP) held constant (A), with pump speed and SP held constant (B) and with pump speed and ST held constant (C). [Hydraulic retention time, HRT].
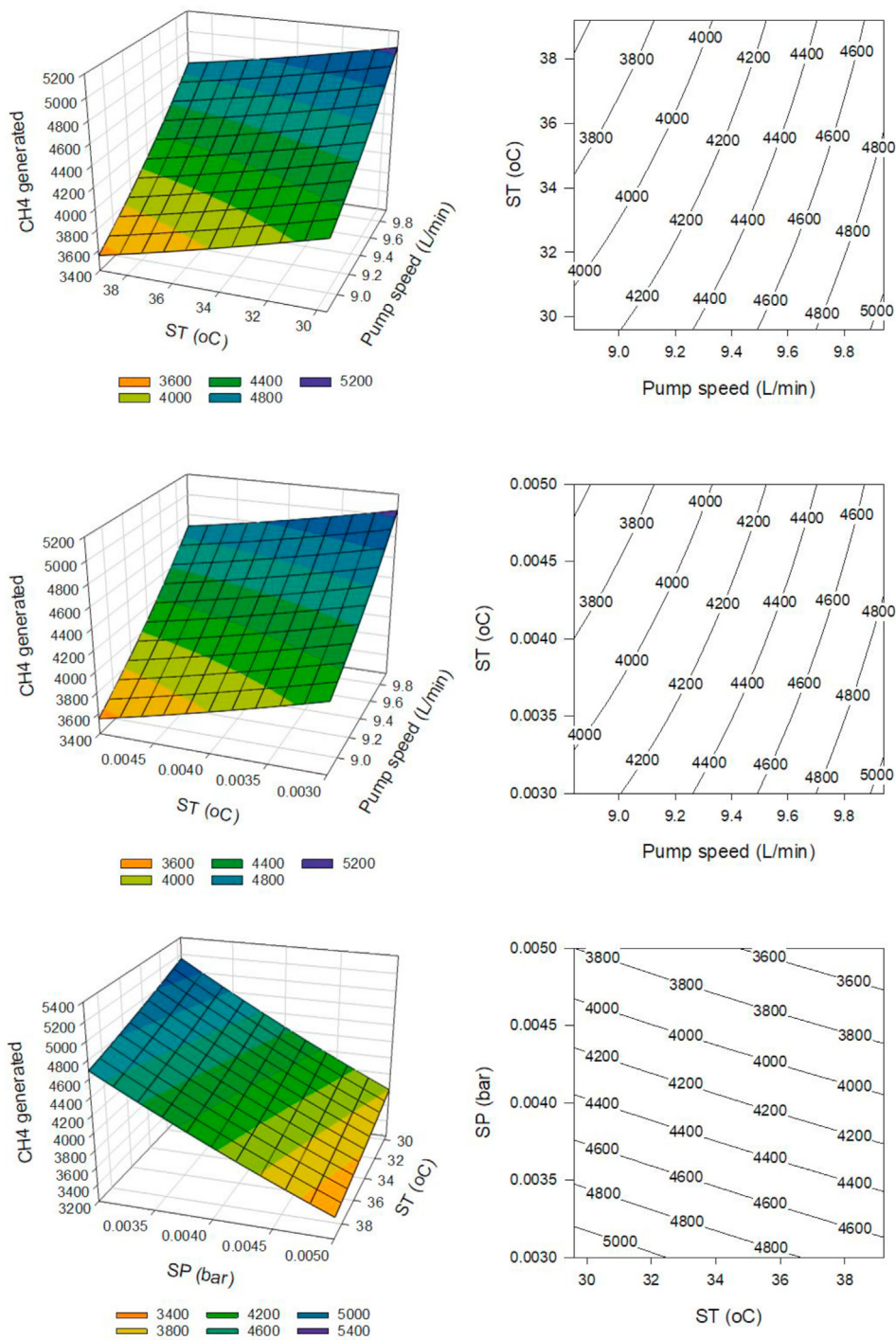
Fig. A.3. Surface plots (left) and corresponding contour plots (right) showing the effects of the H²AD process conditions on the total suspended solids (TSS) removal as predicted by the ECR-ANN model with hydraulic retention time (HRT) and system pressure (SP) held constant (A), with HRT and set temperature (ST) held constant (B) and with HRT and pump speed (C).

Fig. A.4. Surface plots (left) and corresponding contour plots (right) showing the effects of the H^2AD process conditions on the methane (CH_4) generated as predicted by the ECR-ANN model with set temperature (ST) and system pressure (SP) held constant (A), with pump speed and SP held constant (B) and with pump speed and ST held constant (C). [Hydraulic retention time, HRT].

Fig. A.5. Surface plots (left) and corresponding contour plots (right) showing the effects of the H^{2}AD process conditions on the methane (CH$_4$) generated as predicted by the ECR-ANN model with hydraulic retention time (HRT) and system pressure (SP) held constant (A), with HRT and set temperature (ST) held constant (B) and with HRT and pump speed (C).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclepro.2021.126242.

References

Abbas, M.H., Norman, R., Charles, A., 2018. Neural network modelling of high pressure CO 2 corrosion in pipeline steels. Process Saf. Environ. Protect. 119, 36–45. https://doi.org/10.1016/j.psep.2018.07.006.

Ahmad, M.W., Mourshed, M., Rezgui, Y., 2017. Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy Build. 147, 77–89. https://doi.org/10.1016/J.ENBUILD.2017.04.038.

Al-Fattah, S.M., Al-Naim, H.A., Al-Naim, H.A., 2009. Artificial-intelligence technology predicts relative permeability of giant carbonate reservoirs. SPE Reservoir Eval. Eng. 12, 96–108. https://doi.org/10.2118/109018-PA.

Al-Sharrah, G., Elkamel, A., Almanssoor, A., 2010. Sustainability indicators for decision-making and optimisation in the process industry: the case of the petrochemical industry. Chem. Eng. Sci. 65, 1452–1461. https://doi.org/10.1016/j.ces.2009.10.015.

Albuquerque, G., Lowe, T., Magnor, M., 2011. Synthetic generation of high-dimensional datasets. IEEE Trans. Visual. Comput. Graph. 17, 2317–2324. https://doi.org/10.1109/TVCG.2011.237.

Baghban, A., Kahani, M., Nazari, M.A., Ahmadi, M.H., Yan, W.-M., 2019. Sensitivity analysis and application of machine learning methods to predict the heat transfer performance of CNT/water nanofluid flows through coils. Int. J. Heat Mass Tran. 128, 825–835. https://doi.org/10.1016/J.IJHEATMASSTRANSFER.2018.09.041.

Bajracharya, Suman, Sharma, Mohita, Mohanakrishna, Gunda, Dominguez Benneton, Xochitl, Strik, David P.B.T.B., Sarma, Priyangshu M., Pant, Deepak, 2016. An overview on emerging bioelectrochemical systems (BESs): technology for sustainable electricity, waste remediation, resource recovery, chemical production and beyond. Renew. Energy 98, 153–170. https://doi.org/10.1016/j.renene.2016.03.002.

Bakshi, B.R., Fiksel, J., 2003. The quest for sustainability: challenges for process systems engineering. AIChE J. 49, 1350–1358. https://doi.org/10.1002/aic.690490602.

Barbon Junior, S., Mastelini, S.M., Barbon, A.P.A.C., Barbin, D.F., Calvini, R., Lopes, J.F., Ulrici, A., 2020. Multi-target prediction of wheat flour quality parameters with near infrared spectroscopy. Inf. Process. Agric. 7, 342–354. https://doi.org/10.1016/j.inpa.2019.07.001.

Bauer, A., Leonhartsberger, C., Bösch, P., Amon, B., Friedl, A., Amon, T., 2010. Analysis of methane yields from energy crops and agricultural by-products and estimation of energy potential from sustainable crop rotation systems in EU-27. Clean Technol. Environ. Pol. 12, 153–161. https://doi.org/10.1007/s10098-009-0236-1.

Baxter, J., 1995. Jonathan, Learning internal representations. In: '. Proc. Eighth Annu. Conf. Comput. Learn. Theory - COLT, vol. 95. ACM Press, New York, New York, USA, pp. 311–320. https://doi.org/10.1145/225298.225336.

Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.

Borchani, H., Varando, G., Bielza, C., Larrañaga, P., 2015. A survey on multi-output regression. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 5, 216–233. https://doi.org/10.1002/widm.1157.

Brissette, F.P., Khalili, M., Leconte, R., 2007. Efficient stochastic generation of multi-site synthetic precipitation data. J. Hydrol. 345, 121–133. https://doi.org/10.1016/J.JHYDROL.2007.06.035.

Burnham, A.J., MacGregor, J.F., Viveros, R., 1999. Latent variable multivariate regression modeling. Chemometr. Intell. Lab. Syst. 48, 167–180. https://doi.org/10.1016/S0169-7439(99)00018-0.

Caruana, R., 1995. Learning many related tasks at the same time with backpropagation. In: Tesauro, G., Touretzky, D.S., Leen, T.K. (Eds.), Adv. Neural Inf. Process. Syst. 7. MIT Press, pp. 657–664. http://papers.nips.cc/paper/959-learning-many-related-tasks-at-the-same-time-with-backpropagation.pdf.

Charte, F., Romero, I., Pérez-Godoy, M.D., Rivera, A.J., Castro, E., 2017. Comparative analysis of data mining and response surface methodology predictive models for enzymatic hydrolysis of pretreated olive tree biomass. Comput. Chem. Eng. https://doi.org/10.1016/j.compchemeng.2017.02.008.

Coley, C.W., Green, W.H., Jensen, K.F., 2018. Machine learning in computer-aided synthesis planning. Acc. Chem. Res. 51, 1281–1289. https://doi.org/10.1021/acs.accounts.8b00087.

Curteanu, S., Piuleac, C.G., Godini, K., Azaryan, G., 2011. Modeling of electrolysis process in wastewater treatment using different types of neural networks. Chem. Eng. J. https://doi.org/10.1016/j.cej.2011.05.104.

Dalmau, M., Atanasova, N., Gabarrón, S., Rodriguez-Roda, I., Comas, J., 2015. Comparison of a deterministic and a data driven model to describe MBR fouling. Chem. Eng. J. https://doi.org/10.1016/j.cej.2014.09.003.

Escrig, J., Woolley, E., Rangappa, S., Simeone, A., Watson, N.J., 2019. Clean-in-place monitoring of different food fouling materials using ultrasonic measurements. Food Contr. 104, 358–366. https://doi.org/10.1016/J.FOODCONT.2019.05.013.

Fisher, O., Watson, N., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2018. Cloud manufacturing as a sustainable process manufacturing route. J. Manuf. Syst. 47

https://doi.org/10.1016/j.jmsy.2018.03.005.

Fisher, O., Watson, N., Escrig, J., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2020. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. Comput. Chem. Eng. 140, 106881 https://doi.org/10.1016/j.compchemeng.2020.106881.

Fogler, H.S., 2017. Essentials of Chemical Reaction Engineering, second ed. Prentice Hall.

H2Ad: The Chemistry, (n.d.). http://h2ad.org.uk/the-chemistry/(accessed December 5, 2020).

Hadavandi, E., Shahrabi, J., Shamshirband, S., 2015. A novel Boosted-neural network ensemble for modeling multi-target regression problems. Eng. Appl. Artif. Intell. 45, 204–219. https://doi.org/10.1016/J.ENGAPPAI.2015.06.022.

Halder, G., Dhawane, S., Barai, P.K., Das, A., 2015. Optimizing chromium (VI) adsorption onto superheated steam activated granular carbon through response surface methodology and artificial neural network. Environ. Prog. Sustain. Energy 34, 638–647. https://doi.org/10.1002/ep.12028.

Hamid, H.A., Jenidi, Y., Somerfield, C., Gomes, R.L., 2016. Predicting the capability of carboxylated cellulose nanowhiskers for the remediation of copper from water using response surface methodology (RSM) and artificial neural network (ANN) models. Ind. Crop. Prod. 93, 108–120. https://doi.org/10.1016/J.INDCROP.2016.05.035.

Himmelblau, D.M., 2000. Applications of artificial neural networks in chemical engineering. Kor. J. Chem. Eng. 17, 373–392. https://doi.org/10.1007/BF02706848.

Hobæk Haff, I., Aas, K., Frigessi, A., 2010. On the simplified pair-copula construction - simply useful or too simplistic? J. Multivariate Anal. 101, 1296–1310. https://doi.org/10.1016/j.jmva.2009.12.001.

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.

Jeong, M., Lee, G.G., 2009. Multi-domain spoken language understanding with transfer learning. Speech Commun. 51, 412–424. https://doi.org/10.1016/j.specom.2009.01.001.

Kabongo, J.D., 2013. Waste valorization. In: Idowu, S.O., Capaldi, N., Zu, L, Das Gupta, A. (Eds.), Encycl. Corp. Soc. Responsib. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 2701–2706. https://doi.org/10.1007/978-3-642-28036-8_680.

Kim, S., 2017. MATLAB Deep Learning with Machine Learning, Neural Networks and Artificial Intelligence, first ed. Apress. https://doi.org/10.1007/978-1-4842-2845-6.

Kocev, D., Ceci, M., 2015. In: Ensembles of Xxtremely Randomized Trees for Multi-Target Regression. Springer, Cham, pp. 86–100. https://doi.org/10.1007/978-3-319-24282-8_9.

Kocev, D., Vens, C., Struyf, J., Džeroski, S., 2007. Ensembles of multi-objective decision trees. In: Mach. Learn. ECML 2007. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 624–631. https://doi.org/10.1007/978-3-540-74958-5_61.

Kocev, D., Džeroski, S., White, M.D., Newell, G.R., Griffioen, P., 2009. Using single-and multi-target regression trees and ensembles to model a compound index of vegetation condition. Ecol. Model. 220, 1159–1168. https://doi.org/10.1016/J.ECOLMODEL.2009.01.037.

Kolouri, S., Park, S.R., Thorpe, M., Slepcev, D., Rohde, G.K., 2017. Optimal mass transport: signal processing and machine-learning applications. IEEE Signal Process. Mag. 34, 43–59. https://doi.org/10.1109/MSP.2017.2695801.

Larson Lesnik, K., Liu, H., 2017. Predicting Microbial Fuel Cell Biofilm Communities and Bioreactor Performance Using Artificial Neural Networks. https://doi.org/10.1021/acs.est.7b01413.

Levatić, J., Ceci, M., Kocev, D., Džeroski, S., 2017. Self-training for multi-target regression with tree ensembles. Knowl. Base Syst. 123, 41–60. https://doi.org/10.1016/J.KNOSYS.2017.02.014.

Ling, J., Jones, R., Templeton, J., 2016. Machine learning strategies for systems with invariance properties. J. Comput. Phys. 318, 22–35. https://doi.org/10.1016/J.JCP.2016.05.003.

Liu, H., Cocea, M., 2017. Semi-random partitioning of data into training and test sets in granular computing context. Granul. Comput. 2, 357–386. https://doi.org/10.1007/s41066-017-0049-2.

Liu, Q., Xu, Q., Zheng, V.W., Xue, H., Cao, Z., Yang, Q., 2010. Multi-task learning for cross-platform siRNA efficacy prediction: an in-silico study. BMC Bioinf. 11, 181. https://doi.org/10.1186/1471-2105-11-181.

Liu, Y., Pan, Y., Huang, D., Wang, Q., 2017. Fault prognosis of filamentous sludge bulking using an enhanced multi-output Gaussian processes regression. Contr. Eng. Pract. 62, 46–54. https://doi.org/10.1016/J.CONENGPRAC.2017.02.003.

Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., Kallel, A., 2020. A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection. Sci. Total Environ. 715, 136991. https://doi.org/10.1016/j.scitotenv.2020.136991.

Melki, G., Cano, A., Kecman, V., Ventura, S., 2017. Multi-target support vector regression via correlation regressor chains. Inf. Sci. (Ny) 415–416, 53–69. https://doi.org/10.1016/J.INS.2017.06.017.

Min, Q., Lu, Y., Liu, Z., Su, C., Wang, B., 2019. Machine learning based Digital Twin framework for production optimization in petrochemical industry. Int. J. Inf. Manag. 49, 502–519. https://doi.org/10.1016/j.ijinfomgt.2019.05.020.

OECD, 2012. OECD Environmental Outlook to 2050. OECD Publishing. https://doi.org/10.1787/9789264122246-en.

Panerati, J., Schnellmann, M.A., Patience, C., Beltrame, G., Patience, G.S., 2019. Experimental methods in chemical engineering: artificial neural networks—ANNs. Can. J. Chem. Eng. 97, 2372–2382. https://doi.org/10.1002/cjce.23507.

Ranjan, R., Patel, V.M., Chellappa, R., 2019. HyperFace: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. IEEE Trans. Pattern Anal. Mach. Intell. 41, 121–135. https://doi.org/10.1109/TPAMI.2017.2781233.

Saad, M.H., Nazzal, M.A., Darras, B.M., 2019. A general framework for sustainability assessment of manufacturing processes. Ecol. Indicat. 97, 211–224. https://doi.org/10.1016/j.ecolind.2018.09.062.

Shahin, S.A., Mossad, M., Fouad, M., 2019. Evaluation of copper removal efficiency using water treatment sludge. Water Sci. Eng. 12, 37–44. https://doi.org/10.1016/j.wse.2019.04.001.

Spyromitros-Xioufis, E., Tsoumakas, Grigorios, Groves, W., Vlahavas, I., Tsoumakas, G., 2016. Multi-target regression via input space expansion: treating targets as inputs. Mach. Learn. 104, 55–98. https://doi.org/10.1007/s10994-016-5546-z.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.

Stock, T., Seliger, G., 2016. Opportunities of sustainable manufacturing in Industry 4.0. Procedia CIRP 40, 536–541. https://doi.org/10.1016/j.procir.2016.01.129.

Te Liew, W., Adhitya, A., Srinivasan, R., 2014. Sustainability trends in the process industries: a text mining-based analysis. Comput. Ind. 65, 393–400. https://doi.org/10.1016/j.compind.2014.01.004.

Tools: Water Efficiency in the Food and Drink industry | WRAP UK, (n.d.). http://www.wrap.org.uk/node/31479 (accessed August 13, 2019).

Tsoumakas, G., Spyromitros-Xioufis, E., Vrekou, A., Vlahavas, I., 2014. In: Multi-target Regression via Random Linear Target Combinations. Springer, Berlin, Heidelberg, pp. 225–240. https://doi.org/10.1007/978-3-662-44845-8_15.

UK Power, Compare Energy Prices Per kWh | Gas & Electric Per Unit | UKPower, (n.d.). https://www.ukpower.co.uk/home_energy/tariffs-per-unit-kwh (accessed September 30, 2019).

Ulissi, Z.W., Medford, A.J., Bligaard, T., Nørskov, J.K., 2017. To address surface reaction network complexity using scaling relations machine learning and DFT calculations. Nat. Commun. 8, 14621. https://doi.org/10.1038/ncomms14621.

U.S. Energy Information Administration, 2016. International Energy Outlook 2016. Washington. www.eia.gov. (Accessed 2 September 2019).

Wang, J., Ma, Y., Zhang, L., Gao, R.X., 2018. Deep learning for smart manufacturing: methods and applications. J. Manuf. Syst. 48, 144–156. https://doi.org/10.1016/J.JMSY.2018.01.003.

Xiong, T., Bao, Y., Hu, Z., 2014. Multiple-output support vector regression with a firefly algorithm for interval-valued stock price index forecasting. Knowl. Base Syst. 55, 87–100. https://doi.org/10.1016/J.KNOSYS.2013.10.012.

Yildiz, B., Bilbao, J.I., Sproul, A.B., 2017. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew. Sustain. Energy Rev. 73, 1104–1122. https://doi.org/10.1016/j.rser.2017.02.023.

Zarte, M., Pechmann, A., Nunes, I.L., 2019. Decision support systems for sustainable manufacturing surrounding the product and production life cycle – a literature review. J. Clean. Prod. 219, 336–349. https://doi.org/10.1016/j.jclepro.2019.02.092.