SOCAR TÜRKİYE R&D

Journal Publication

SOCAR
AR-GE

2023

Contents

Introduction

The SOCAR R&D Journal is a collection of Academic works that were published in the years 2021-2023 by the R&D researchers. Committed to its 51-year research, remarkable development and constant innovation, SOCAR R&D and Innovation Company in line with SOCAR Global and SOCAR Türkiye aim to create an integrated energy business that adds value to all stakeholders in an economically and socially responsible way. SOCAR Türkiye R&D fields of expertise include catalyst research, green chemistry, sustainable, clean, and efficient energy, value-added products, energy transition, digitalization, and environmental solutions. Adoption of an open innovation model allows the SOCAR R&D to engage in strong industry-academia collaborations and work with more than 30 national and international universities. SOCAR R&D currently holds 9 patents and 18 applications for national and EU patent.

Regarding SOCAR R&D ongoing Project activities, the SOCAR R&D Journal consists of a total of 8 articles classified into 3 groups as Process Digitalization (6 articles), Sustainability & Catalyst (1 article), and Product & Biotechnology (1 article). These articles have been published in high-impact factor journals, such as *Computers & Chemical Engineering Journal, Computer Aided Chemical Engineering, Processes, Polymers and Archives of Microbiology*. Most of these articles have been presented at different national and international conferences such as the 31st European Symposium on Computer Aided Process Engineering (Istanbul, Türkiye), European Symposium on Computer-Aided Process Engineering-32 (Toulouse, France), the 9th International Conference on the Foundations of Systems Biology in Engineering (Boston, USA), etc. The research articles have been written in academic collaboration with professors at METU, Boğaziçi, Ege, and other national universities.

1
Optimal Artificial Neural Network Architecture Design for Modeling an Industrial Ethylene Oxide Plant

Hasan Sildir[a*], Sahin Sarrafi[b], Erdal Aydin[c,d,e]

[a]*Department of Chemical Engineering, Gebze Technical University, Kocaeli 41400, Türkiye*
[b]*SOCAR Türkiye, R&D and Innovation Co., Aliaga, Izmir, 35800, Türkiye*
[c]*Department of Chemical Engineering, Bogazici University, Bebek, İstanbul 34342, Türkiye*
[d]*Department of Chemical and Biological Engineering, Koç University, Istanbul 34457, Türkiye*
[e]*Koç University TUPRAS Energy Center (KUTEM), Koç University, Istanbul, 34450, Türkiye*
[*]*hasansildir@gtu.edu.tr*

Journal: Computers & Chemical Engineering Journal, Volume 163, **2022**, 107850.

https://www.sciencedirect.com/science/article/abs/pii/S0098135422001880

DOI: https://doi.org/10.1016/j.compchemeng.2022.107850

Abstract

Optimum selection of input variables, number of hidden neurons and connections among the network elements deliver the best configuration of an ANN, usually resulting in reduced over-fitting and improved test performance. This study focuses on the development of a superstructure-oriented feedforward ANN design and training algorithm whose impacts are demonstrated on an industrial Ethylene Oxide (EO) plant for the prediction of product related variables. Proposed method brings about a mixed integer nonlinear programming problem (MINLP) to be solved, which takes the existence of inputs, neurons, and connections among the network elements into account by binary variables in addition to continuous weights of existing connections. Our investigations show that almost 85% of the ANN connections are removed compared to the fully connected ANN (FC-ANN) with 50% decrease in the number of inputs of the ANN. The modified ANN delivers a better prediction performance over FC-ANN, since FC-ANN suffers from over-fitting.

Keywords: machine learning; artificial neural networks; superstructure optimization; process modelling; mixed integer nonlinear programming.

Introduction

Artificial Neural Network (ANN) is a sophisticated integration of high number of relatively simple mathematical expressions (activation functions) and input variables to deliver an output prediction based on network parameters and architecture. ANNs involve significant number of connections, further characterized by parameters, among network elements having the capability

to represent a broad range of data. Various types of ANNs have been developed (Gupta and Raza, 2019) after the foundations were laid earlier (McCulloch and Pitts, 1943). Feedforward ANNs propagate the information in the network only in the forward direction through the output. The parameters which are required for the information propagation, in addition to network architecture, are obtained using historical measurements; thus, the method is exposed to typical weaknesses of data driven empirical models.

Implementations of ANNs cover a wide spectrum of research areas. Okuyucu et al (Okuyucu et al., 2007) used ANNs to calculate the correlations between the friction stir welding parameters of aluminum and its mechanical properties. In (Kalyan et al., 2014), ultrasound images of liver are considered and disease conditions were calculated using a classification objective function. ANNs found potential also in pharmaceutical industry (Agatonovic-Kustrin and Beresford, 2000). Hossain used twelve air related inputs to calculate the dispersion of the pollutants over different regions of urban conditions (Hossain, 2014). A detailed review on the implications for process systems engineering, in particular, is provided in (Lee et al., 2018). A more focused approach on chemical engineering can be found in (Venkatasubramanian, 2019). A comprehensive survey of ANNs applications in a wider vision are given in (Abiodun et al., 2018).

The standard application of ANNs includes mostly fully connected networks, where all inputs, neurons and outputs are entirely connected. Fully connected ANN architectures (FC-ANNs) usually comprise high number of parameters. Typically, as the dimensions get larger, higher number of connections and parameters are introduced. As a result, it is usually expected that the increase in the number of parameters provides higher capability of fitting to the training data. Nevertheless, it is worthwhile here to state that this may easily result in overfitting, and therefore poor prediction capability of the ANN. Training of ANNs is a challenging task due to nonconvexity issues, leading to suboptimal and different solutions based on the initial guess of nonlinear optimization problem (NLP) solved for training (Schweidtmann and Mitsos, 2019). In addition, significant theoretical advancements are needed to ensure global optimality (Haeffele and Vidal, 2017). Such problems arise more often once the overall architecture gets larger, in general. On the other hand, for the training in a traditional sense, the architecture must be prespecified, leaving only the continuous parameters as decision variables.

The hyper-parameter related issues including the selection of inputs, the determination of the number of hidden layer neurons and the connectivity within the network are crucial as they have significant impact on the overfitting problem, in addition to computational and practical issues. Such parameters are also exposed to identification issues with statistically or spatially limited measurements (Levasseur et al., 2017). Introducing more data is usually

not a satisfactory effort as new data might not carry additional statistical information unless they are collected from a different location, which might be challenging task especially for industrial plants.

There are Bayesian optimization-based methods available in the literature for ANN hyper-parameter optimization problem, in particular for the feature selection. Yet, such derivative-free methods might result in different local solutions and the global solution cannot be assured. Moreover, Bayesian optimization is currently non-parallelizable. On the other hand, efficient decomposition algorithms are present for larger search spaces, integrating the derivate-based and blackbox optimization approaches for the solution of MINLPs (Diaz et al., 2017; Feurer and Hutter, 2019; Stamoulis et al., 2018).

Another typical solution to overfitting reduction problem is to include a regularization term to penalize larger values of ANN parameters during the NLP solution, which unfortunately cannot regularize the hyper parameters (Manngård et al., 2018). Dropout regularization is an alternative and efficient method to include structure detection element into ANNs, using random sampling based techniques (Poernomo and Kang, 2018).

Similarly, pruning is another method to reduce the number of connections in ANNs (Xie et al., 2019; Zhou et al., 2019). It is a method to reduce the number of connections in ANNs (Akyol, 2020; He et al., 2019; Sietsma and Dow, 1988; Xie et al., 2019). Kavzoglu and Mather (Kavzoglu and Mather, 1999) compared Magnitude-based pruning (Hassibi and Stork, 1993), Optimum brain damage (Le, 1990) and optimal brain surgeon (Hassibi and Stork, 1993) for the land cover classification. Those methods eliminate some of the connections based on the sensitivities and second order derivatives, followed by the re-training of the reduced network. It was shown in (Kavzoglu and Mather, 1998) that, the accuracy might increase to some extent when some connections are eliminated. On the other hand, these methods usually require many sequential steps and cannot be considered as automatic and simultaneous methods.

Selection of the related features from a comprehensive data is a challenging but critical problem since high number of inputs introduce more parameters into the mathematical ANN formulation. More concretely, selecting the best features to be employed both for offline and online applications is an inevitable task to fully integrate the benefits of machine. A detailed review on feature selection is provided in (Miao and Niu, 2016) and various advantages are provided in (Kavzoglu and Mather, 2002). Moreover, Verikas and Bacauskiene introduced an augmented cross-entropy error function for the classification problems using ANNs (Verikas and Bacauskiene, 2002) for the feature selection. Jensen et al. used Fisher Discrimination for the selection of ANN features (Jensen et al., 2001).

Several papers are present in the literature discussing training, MILP optimization and verification for neural networks using ReLU activation functions. Nevertheless, in this work, we discuss the nonlinear activation function case using the hyperbolic tangent (or a similar) to address the increased nonconvexity issues due to high nonlinearity in the hidden layer. Finally, proposed method not only detects the optimal features but also the number of neurons and the optimal connection configuration between them (Anderson et al., 2020; Bunel et al., 2020; Grimstad and Andersson, 2019; Lodi and Nagarajan, 2019; Rössig and Petkovic, 2021).

In a similar fashion, Dua proposed using a general mixed-integer optimization formulation to eliminate overfitting by detecting the optimal configuration of ANNs (Dua, 2010). Both number of neurons and existence of the interconnections are included in the objective function. Yet, this formulation does not consider the selection of optimum input variables. Moreover, resulting formulations are either mixed-integer linear programs with fixed parameter weights or small scale MINLPs with fixed structures, number of nodes and/or interconnections, exhibiting poor flexibility. On the other hand, this significant study shows that much better test, and thus prediction performance can be achieved with fewer neurons and connections.

In this study, a novel MINLP (mixed-integer nonlinear programming) formulation is developed for the simultaneous design and training of an optimal architecture feedforward ANN (OA-ANN). This method requires the modification of traditional ANN equations. Suggested method not only accounts for the optimal features to be selected, but also for the optimal neuron number in the hidden layer and configuration of the connections among network variables. As a result, the optimal topology design problem can be simultaneously integrated with the training problem with a more flexible structure detection and efficient reduction in input space, in addition to significant reduction in network connections. Moreover, please note that even though there are some widely used methods to reduce overfitting using cross validation, these methods usually require several sequential trial and fitting steps to detect the optimal number of neurons and features. Conversely, the proposed MINLP-based method realizes the steps in a simultaneous and automatic fashion. Optimum structure, features and connections are obtained after solving only one optimization problem, which is flexible and fine-tuned by further modifications in the formulation based on problem-specific needs or computational power. A heuristic solution approach is developed for the highly non-convex MINLP formulation with significant number of both discrete and continuous decision variables. The details of the approach are provided in the methodology section. Next, the implementation on industrial plant is presented. Last section concludes the study.

Process Description

Ethylene oxide (EO), a captive feedstock for ethylene glycols (MEG, DEG, and TEG), is produced through selective oxidation of ethylene and oxygen in the presence of an Ag supported α-Al_2O_3 catalyst. The EO process, in SOCAR Türkiye Petkim Petrochemical plant, takes place in two parallel fixed-bed multi-tubular reactors at a temperature and pressure range of 240–260 °C and 17–18 bar, respectively. Both ethylene epoxidation and EO combustion occur in the reactors and only the former is desirable. Side reactions reduce the EO yield and an increase in CO_2 emission from the plant. From the point of both safety and optimum control perspectives, the monitoring of the concentrations of the streams, which are measured through two separate online gas chromatography (GC) instruments, is of great importance. The former is designed for hydrocarbons and permanent gases whereas the latter focuses on the chlorine-based compounds. The operating strategy of the reactors is to maximize the EO yield under various operational constraints. The simplified process flow diagram of the process is shown in Fig. 1.

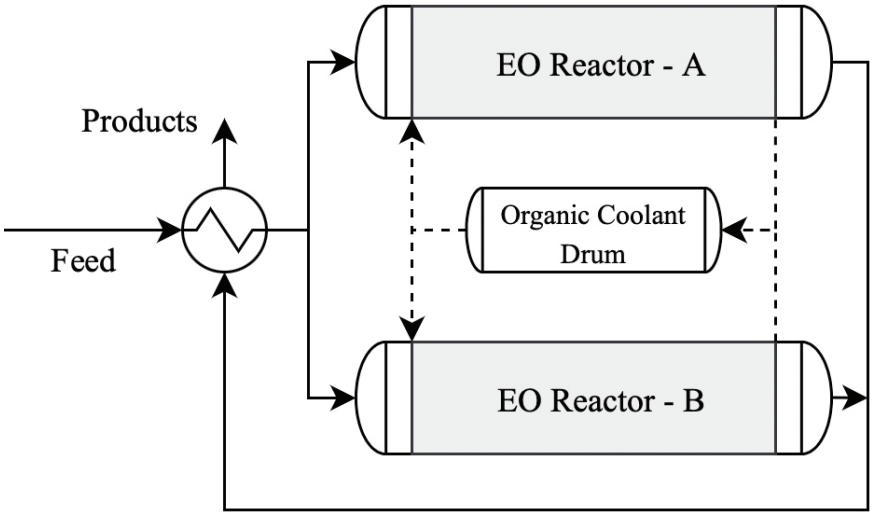

Figure 1. Simplified process flow diagram of EO process

With aging of the catalyst, the trade-off between selectivity and productivity becomes gradually more prominent. Along with the compensation of loss of active sites under the favor of increasing of the temperature, catalyst still provides sustainable commercial yields at the expense of excess feedstock consumption which in turns leads to boost GHG emissions by releasing more carbon dioxide (CO2) into the atmosphere. To maintain catalyst activity for a longest period possible, controlling process variables more preciously with a robust model is very demanding issue throughout the last two decades.

Many criteria must be considered during EO production at the desired levels under the continuous and sluggish catalyst deactivation. Related information is obtained using the features given in Table 1. All the feed compositions are measured before the reactor (i.e., after EO and CO_2 absorption section). Fresh VCM feed is introduced into the reactor inlet stream right before the reactor. Real-time optimization and automation of such a complex process is a challenging task, requiring significant effort in first principles and mechanistic modeling. Alternatively, Artificial Neural Networks (ANNs) are promising empirical models to estimate the product related variables from easily measurable variables (i.e. temperatures and pressures).

Table 1. Descriptions of the features and inputs

TAG Description	Unit	TAG Description	Unit
Lean absorbent flow rate	t/h	Fresh C_2H_4 feed	t/h
CO_2 absorber top pressure	kg/cm^2	Fresh CH_4 feed	kg/h
Feed composition (C_2H_4)	% (v/v)	CH_3Cl (methyl chloride) in the recycle	ppm
Feed composition (O_2)	% (v/v)	VCM (Viniyl chloride monomer) in the recycle	ppm
Feed composition (N_2)	% (v/v)	C_2H_5Cl (ethyl chloride) in the recycle	ppm
Feed composition (Ar)	% (v/v)	Reactor A coolant pressure	kg/ cm^{2}g
Feed composition (CH_4)	% (v/v)	Reactor A inlet pressure	kg/ cm^{2}g
Feed composition (C_2H_6)	% (v/v)	Flow rate to reactor A	t/h
Feed composition (CO_2)	% (v/v)	Fresh VCM feed to reactor A	kg/h
Feed composition (EO)	% (v/v)	Product composition (EO)	% (v/v)
Fresh O_2 feed	t/h		

Methodology: Optimal Superstructure-based Design of ANNs

A feedforward ANN model is generally expressed as:

$$y = f_1(A \cdot f_2(B \cdot u + C) + D) \quad (1)$$

where f_1 and f_2 are output and hidden layer activation functions respectively. A typical activation function is the hyperbolic tangent function, which is used in this study. A and B are weight matrices; C and D are bias vectors; u is the input vector, and y is the output vector. Related ANN parameters to be calculated in

the training are all continuous, and theoretically unbounded. Moreover, their dimensions depend on the number of inputs, outputs, and number of neurons determined manually before training. As aforementioned before, identification issues of these ANN parameters usually result in poor testing performance.

The MINLP formulation suggested in this work introduces binary variables to the traditional ANN equations to represent the existence of network elements. The resulting superstructure formulation also takes the selection of the input variables (features) into account in addition to the number of hidden neurons and connections, which in turn maximizes the overall flexibility and strength of the formulation. Accordingly, the proposed MINLP problem, whose objective is to minimize the number of connections of the ANN, is given by:

$$\min_{A,A_{binary},B,B_{binary},C,D,N_{binary},U_{binary}} \sum_{i,j} A_{binary,ij} + B_{binary,ij}$$

$$s.t.\quad F = \sum_{i=1}^{N} \left\| f_1\left(\left(A \circ A_{binary}\right) \cdot diag\left(N_{binary}\right) \cdot f_2\left(\left(B \circ B_{binary}\right) \cdot diag\left(U_{binary}\right) \cdot u_i + C\right) + D\right) - y_i \right\| \tag{2}$$

$$F \leq \mathcal{E};$$

$$A_{binary,ij} \leq N_{binary,j};\ B_{binary,ij} \leq U_{binary,j};$$

$$-B_{LB} \times B_{binary,j} \leq B_{i,j} \leq B_{UB} \times B_{binary,j};$$

$$-A_{LB} \times A_{binary,j} \leq A_{i,j} \leq A_{UB} \times A_{binary,j};$$

$$A_{i,j}, B_{i,j}, C, D \in [-4,4];\ A_{binary}, B_{binary}, N_{binary}, U_{binary} \in 0,1$$

where $\circ$ is the Hadamard product operator; u_i and y_i are the i^{th} input and output samples, respectively. N is the number of samples used for the training. $A_{binary,ij}$ and $B_{binary,ij}$ are matrices with binary values representing the existence of connections. N_{binary} and U_{binary} are the binary vectors defining the existence of a neuron and an input, respectively. The existence of a particular connection between a neuron and an input is defined by the binary variable $B_{binary,ij.}$ A_{ij} is the continuous weight parameter of the connection between the j^{th} neuron and the i^{th} output. Similarly, B_{ij} represents the connection between an input and the corresponding neuron. As a result, the optimal input is detected by optimizing the input neurons for a fully connected feedforward network, which also eliminates the corresponding weights from B_{binary} through the linking constraints. Such constraints further tighten the formulation and enable a more efficient search in the optimization as when the corresponding input neuron is not selected, or if it has zero weight, the corresponding input (or feature) is not preferred. Please also note that this formulation is easily expandible for deep neural networks for optimum connection configuration through several hidden layers. $\mathcal{E}$ is the upper bound for the overall training error. This way, a trade-off is included into the training, which in turn is expected to increase the test performance after implementing the proposed formulation. Several trials showed that the upper and lower limits corresponded to 4 and -4, respectively, for the continuous weight values for the *tanh* activation function. In order

words, activation function saturation occurs beyond these limits, in our case. Yet, interval analysis could also be performed to tighten the MINLP problem in a more rigorous way. Finally, please note that suggested formulation has the flexibility to result in a fully connected network once ε is set to a small value. In other words, a fully connected ANN is the upper bound of the network obtained using the suggested formulation (OA-ANN).

Problem (2) displays highly nonlinear and non-convex behavior not only due to the presence of the nonlinear activation functions in layers but also because of the discrete decision variables. There is notable number of papers in the literature addressing the solution of MINLP problems (Boukouvala et al., 2016; Kocis and Grossmann, 1988; Kronqvist et al., 2019; Schlueter, 2014). Some popular and efficient ways for solving MINLPs are discrete branch and bound methods (Lawler and Wood, 1966), generalized Benders decomposition (Geoffrion, 1972) and outer approximation (Duran and Grossmann, 1986). However, please note that these methods are proven to be efficient for convex problems; do not guarantee global convergence for non-convex problems and usually diverge unless being used with reformulation techniques (Smith and Pantelides, 1999). Commercial codes, employing special algorithms and reformulation/relaxation techniques for solving non-convex MINLPs to global optimality (Kesavan et al., 2004; Misener and Floudas, 2014; Sahinidis, 1996), are also available. These methods often decompose the original non-convex MINLP into mixed-integer linear programs (MILPs) and nonlinear programming problems (NLPs). As a result, the MILP can be solved via an efficient branch and bound, while the NLP is solved using a global nonlinear optimization algorithm. Unless a global algorithm is used for the decomposed NLP, the overall solution procedure might result in an infeasible point or even diverge. On the other hand, note that one vital drawback of the derivative based global methods is the required computational load.

Another alternative for solving non-convex MINLPs is to use derivative-free evolutionary algorithms or to employ heuristic methods (Bonami and Gonçalves, 2012). Similar to rigorous methods, heuristic methods suggest decomposing the original non-convex MINLP. Yet, unlike rigorous and derivative based global optimization algorithms, the resulting NLP is solved by a local optimization method, which has proven to be efficient for certain type of applications (Pintarič and Kravanja, 2000; Wen and Ma, 2008).

In our case, the approach in (Kocis and Grossmann, 1988) was tested to solve Problem (2) and performed poorly due to the high non-linearity stemming from the activation functions as mentioned above. Further relaxation techniques might be employed, and the application of the aforementioned derivative free global methods may result in much better convergence performance (Joy et al., 2019; Schweidtmann and Mitsos, 2019). Yet, employing derivative based global algorithms and decomposition techniques in (Smith and Pantelides, 1999) is not in the scope of this paper.

In this work, an adaptive and evolutionary solution algorithm is used to solve the resulting non-convex MINLP problem. This method exhibits a similar idea with (Pintarič and Kravanja, 2000) and it decomposes the original MINLP into an integer program (IP) and a nonlinear program (NLP). IP comprises merely binary decision variables that can be adjusted during optimization, whereas NLP involves continuous decision variables. The IP stands on the outer loop and is solved via the genetic algorithm based IP solver of MATLAB while the inner loop NLP is solved using a rigorous derivative based NLP solver, IPOPT (Wächter and Biegler, 2006). Here, we should discuss the fact that while the inner NLP calculates the optimum weights and biases firstly under constraints specified by the genetic algorithm which optimizes the input variables and the existence of the neurons and connections based on the F^*, the training error based on the particular architecture. Note that F and F^* are equal only when NLP iterations result in a converged training. Nevertheless, the outer loop does not have the full domain knowledge and treats the lower level calculations a blackbox block, and thus we call it as a heuristic method. The implementation is shown in Fig. 2 and is not the most rigorous type neither in terms of dual decomposition nor bilevel programming. However, similar implementations and alternative heuristic MINLP solution approaches have found wide applications in the literature as well (Alipour et al., 2018; Elsido et al., 2017; Evins, 2015; Huang et al., 2019; Jerez et al., 2014). In order to circumvent this issue, more rigorous derivative based global MINLP solution algorithms could be utilized, which would result in global solutions (Grossmann, 2013). Yet, employing these types of algorithms is not in the scope of this work, the main aim is to show the potential of the suggested MINLP-based formulations. Finally, even the proposed heuristic method brings about a better tailored ANN structure compared to the fully connected ANN, which will be discussed in the case study section.

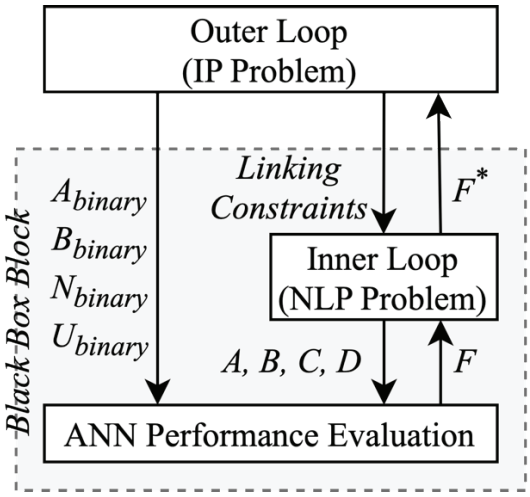

Figure 2. Overall heuristic solution structure

The inner and outer loops are solved sequentially until the tolerance value of the original problem objective value, or the maximum wall clock time is reached. In this study, this criterion is on the change of the original problem objective, with a value of 0.01. This quasi-decomposition approach provided satisfactory performances in our cases. Investigations show that suggested method usually provides a sufficient heuristic solution in less than 3 min. Nevertheless, please note that training optimization problems for ANNs exhibit highly nonlinear and non-convex behavior, calling for many local and suboptimal solution points. Furthermore, suggested iterative solution method has the flexibility to result in a fully connected network.

Even though Problem (2) is generally a non-convex and large-scale MINLP problem, also quite challenging to solve to the global optimum, in show that suggested method usually provides a sufficient heuristic solution in less than 3 min. On the other hand, training ANNs with derivative free solvers is a common subject (Kim and Han, 2000).

Finally, please note that IPOPT is a local solver and might add randomness to the outer problem as the IP solution is based on a black-box genetic algorithm. Thus, we do not propose any integer cuts to be provided during the iterations. In addition to that, sensitivity analysis of the proposed method might be of interest to fully assess the regularization of the optimal ANNs using heuristic solutions.

Case study on the Industrial Ethylene Oxide Plant

The proposed MINLP based approach is tested on the industrial EO plant using hourly measurements of 105 randomly selected days at various training ratios. The suggested MINLP problem has 221 continuous, including bias weights, and 240 discrete variables.

Although the main focus of the proposed MINLP-based approach is in artificial neural networks, preliminary machine learning method are compared in order to justify the usage of ANN methods for the ethylene oxide plant. Initial studies on the soft sensor development for the plant show that artificial neural network and decision tree methods deliver the best performance in terms of testing instances. Detailed training and test performance comparison of different machine learning approaches at 80% training ratio are presented in Table 2a for mean squared error. All results are obtained using Sklearn toolbox of Python using 5-fold cross validation. Multiple linear regression and random forest results are reported by subtracting the large outliers from both train and test data set. Otherwise, very large values are obtained, which further illustrates that these methods are not preferable. Since both fully-connected and dropout neural networks delivered the best performance, it is concluded that neural networks are preferable for the investigated plant.

Table 2a. Different machine learning method performance comparisons

Method	MSE	
	Training	Test
Multiple Linear Regression	1.8	2.5
Random Forest[§]	1.3	1.7
Decision Tree	0.0001	0.03
Support Vector Regression[‡]	0.5	0.042
Fully Connected Artificial Neural Network (FC-ANN)[¤]	0.005	0.02
Dropout Artificial Neural Network (DP-ANN)	0.01	0.012

[§]Trained using 500 estimators

[‡]Trained using radial basis functions

[¤]Trained using lbfgs algoritm in 200 epochs. Includes 10 hidden *tanh* neurons.

OA-ANN architecture using 20% of the data in the training is shown in Fig. 3 as a typical demonstration. 9 inputs, out of 20, are selected from which connections exist to hidden layer as shown in Fig. 3 based on the simultaneous design and training proposed by this study. The thickness of the existing connections represents the absolute magnitude of the corresponding weight. In OA-ANN, only significant connections, both in terms of identifiability and prediction accuracy considerations, are maintained, eliminating the ineffective connections.

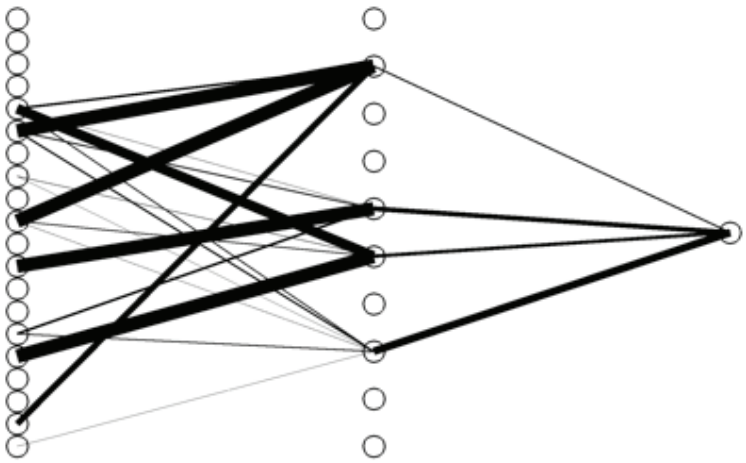

Figure 3. OA-ANN for 20% training data

In addition, the OA-ANN architecture in Fig. 3, with only 4 hidden layer neurons, does not have a fully connected architecture as the selected inputs do not have connection to all existing hidden neurons. As a result, a more efficient information flow is obtained through the network, in which redundant or

statistically insignificant connections are eliminated while still satisfying a desired accuracy in the training.

Selected features include feed composition of N_2, Ar, C_2H_6 and EO, fresh C_2H_4 and VCM feed, VCM, ethyl chloride concentration in the recycle, and flow rate of the reactor A. From process point of view, the chemical components (N_2, Ar an C_2H_6), which is a subset of the selected features by the OA-ANN, act as inert in the reactor occupying a prominent place in the EO yield. Of these, ethane has been engaging into the surface reactions with VCM on the catalyst, and subsequently converting into ethyl chloride. This feature is also critical for the adjustment of selectivity.

Moreover, VCM concentration in the streams, both in recycle and fresh feed, are of central concern due to selectivity and parsimonious control of catalyst life. EO composition in the feed has also an adverse effect on EO yield in the sense that it inhibits the desired reaction, which is partial oxidation of ethylene. Therefore, it is tightly controlled in EO absorber unit operation where the concentration of EO in the recycle is determined and kept minimum as much as possible. Other prominent features are fresh ethylene feed and reactor flow rate. These are the main drivers in the plant production rate, and therefore on the EO yield. Therefore, identified features by OA-ANN have an impact on EO yield from the standpoint of process engineering.

Table 2b. OA-ANN and FC-ANN detailed performance comparison

			OA-ANN		FC-ANN		DP-ANN	
			Train	Test	Train	Test	Train	Test
Training ratio	70%	RMSE	0.0045	0.005	0.0004	0.047	0.004	0.006
		CV	0.065	0.093	0.017	0.296	0.051	0.061
		MAE	-	0.37	-	1.03	-	0.39
	50%	RMSE	0.003	0.0035	0.0004	0.031	0.004	0.045
		CV	0.051	0.061	0.018	0.25	0.062	0.07
		MAE	-	0.49	-	4.53	-	0.6
	20%	RMSE	0.003	0.004	0.0001	0.0140	0.005	0.02
		CV	0.046	0.061	0.008	0.147	0.058	0.09
		MAE	-	0.424	-	2.481	-	3.2

In order to include another comparison, dropout regularized ANN (DP-ANN) comparison is also added in this step. Yet, the main focus of this study is to compare the performance of the proposed method with the fully-connected neural networks. As mentioned earlier, DP-ANN method includes a random generation based algorithm. Detailed training and test performance

comparison of FC-ANN, dropout layer ANN (DP-ANN) and OA-ANN at various training ratios (i.e. 20%, 50% and 70%) are presented in Table 2b including some common statistical measures. Please note that the results of Table 2a and Table 2b are obtained in different instances and using unlike shuffled data for the same dataset. Therefore, ANN results are not the same in the two tables. A maximum of ten neurons are used for all of the ANNs since this number of neurons provides acceptable performance in Table 2a.

Table 2b includes root mean square error (RMSE), coefficient of variation (CV) and maximum absolute error (MAE) as comparison metrics. Due to eliminated over-fitting, training and test performances of OA-ANN show a good agreement at all training ratios. Although FC-ANN demonstrates significantly lower training error with high number of connections, it suffers from larger test error compared to OA-ANN. Similar insights can also be obtained from CVs. Furthermore, it can also be observed that OA-ANN brings about better test performance not only for higher training ratios, but also for smaller ones. At the same time, even though DP-ANN provides better test performance than the FC-ANN and closer metrics to OA-ANN for larger training ratio, it has degrading performance for larger test ratios. This can be claimed to be expected since dropout is not preferable for lower training ratios due to its random nature; OA-ANN is more robust for all training ratios, which in turn shows another potential of the method for the investigated plant. In other words, this situation demonstrates that suggested method is advantageous for applications where limited number of data or measurements are available, which is a vital side-advantage. Accordingly, in all cases, test performance of OA-ANN is superior to FC-ANN despite higher training error.

The architecture comparisons are given in Table 3. The network dimensions of OA-ANN are significantly less than FC-ANN in all training ratios, providing additional computational advantages when real-time ANN update is necessary.

Table 3. OA-ANN and FC-ANN architecture comparison

	OA- ANN			FC-ANN
	20%	50%	70%	-
Neurons	4	3	5	10
Connections	22	19	26	210
Inputs	9	9	8	20

As mentioned earlier, overfitting may cause poor test performance. Both FC-ANN and OA-ANN test results are demonstrated in addition to normalized measurements in Fig. 4. Impact of overfitting can be observed in Figure 4, where

different operation regimes are given with respect to time. Until sample number 400, one can state that both FC-ANN and OA-ANN deliver similar prediction capability and are in good agreement with the actual data. Nevertheless, please observe that the OA-ANN provides much better prediction performance for a significant upset or change in the operating condition, as shown between sample numbers 400-500. This issue stems from high number of connections in FC-ANN, whose weights could not be identified based on the statistical content of the current training data. As a result, the FC-ANN provides worse prediction performance for changing operation window whereas the performance of the OA-ANN does not deviate significantly when the process window changes. Consequently, the OA-ANN is proven to be a better candidate for real-time optimization and control studies, compared to the fully connected ANN model.

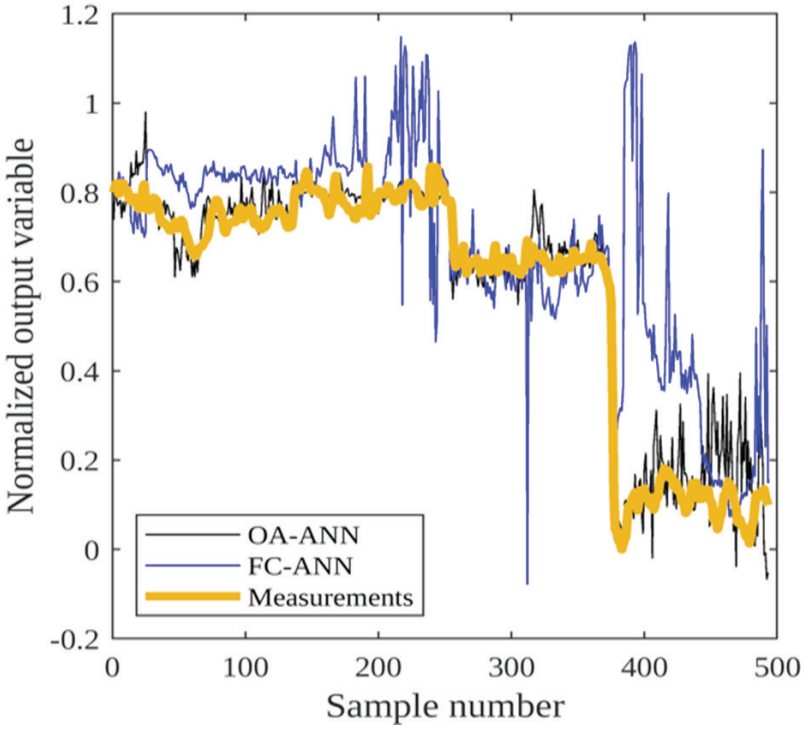

Figure 4. Test data performance of OA-ANN and FC-ANN

Conclusion

Optimal design and structure synthesis for ANNs is a challenging but major task for data-driven real time optimization with ANNs embedded. Ineffectively large networks provide significant test errors due to over-fitting. In this study, an MINLP-based optimal structure detection and feature selection method is proposed. Connection subset selection using an MINLP helps minimizing overfitting effect by eliminating the corresponding weight. As a result, information flow from the corresponding path is eliminated, resulting in reduced correlated or redundant information flow. Moreover, suggested formulation

also takes the existence of neurons and inputs into account in the MINLP as decision variables. Resulting method, called as OA-ANN, provides better estimation capability in various training ratios, more robust the prediction of different operating conditions. In fact, proposed method is more immune to lack of training data due to reduced overfitting. Furthermore, the heuristic MINLP solution algorithm results in a satisfactory network architecture and still has theoretical potential for development.

Suggested method is implemented on an industrial ethylene-oxide plant for real-time optimization purposes. Test performance of the proposed approach (OA-ANN) is significantly better compared to the fully-connected ANN (FC-ANN), despite using approximately 10% of the connections and half of the inputs and neurons only. Reduction in the input space is also advantageous for the real-time operation, when sensor failures are under consideration. In addition, using this method, redundant sensors might be removed from the plant if the soft sensor development is the ultimate functionality.

Acknowledgements

This publication has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C245). However, the entire responsibility of the publication belongs to the owner of the publication.

References

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A.E., Arshad, H., 2018. State-of-the-art in artificial neural network applications: A survey. Heliyon 4, e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. J. Pharm. Biomed. Anal. 22, 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

Akyol, K., 2020. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. Expert Syst. Appl. 140. https://doi.org/10.1016/j.eswa.2019.112875

Alipour, M., Zare, K., Seyedi, H., 2018. A multi-follower bilevel stochastic programming approach for energy management of combined heat and power micro-grids. Energy 149, 135–146. https://doi.org/10.1016/j.energy.2018.02.013

Anderson, R., Huchette, J., Ma, W., Tjandraatmadja, C., Vielma, J.P., 2020. Strong mixed-integer programming formulations for trained neural networks. Math. Program. 183, 3–39. https://doi.org/10.1007/s10107-020-01474-5

Bonami, P., Gonçalves, J.P.M., 2012. Heuristics for convex mixed integer nonlinear programs. Comput. Optim. Appl. 51, 729–747.

Boukouvala, F., Misener, R., Floudas, C.A., 2016. Global optimization advances in Mixed-Integer Nonlinear Programming, MINLP, and Constrained Derivative-Free Optimization, CDFO. Eur. J. Oper. Res. https://doi.org/10.1016/j.ejor.2015.12.018

Bunel, R., Mudigonda, P., Turkaslan, I., Torr, P., Lu, J., Kohli, P., 2020. Branch and bound for piecewise linear neural network verification. J. Mach. Learn. Res. 21.

Diaz, G.I., Fokoue-Nkoutche, A., Nannicini, G., Samulowitz, H., 2017. An effective algorithm for hyperparameter optimization of neural networks. IBM J. Res. Dev. 61, 9:1-9:11. https://doi.org/10.1147/JRD.2017.2709578

Dua, V., 2010. A mixed-integer programming approach for optimal configuration of artificial neural networks. Chem. Eng. Res. Des. 88, 55–60. https://doi.org/10.1016/j.cherd.2009.06.007

Duran, M.A., Grossmann, I.E., 1986. An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. 36, 307–339. https://doi.org/10.1007/BF02592064

Elsido, C., Bischi, A., Silva, P., Martelli, E., 2017. Two-stage MINLP algorithm for the optimal synthesis and design of networks of CHP units. Energy 121, 403–426. https://doi.org/10.1016/j.energy.2017.01.014

Evins, R., 2015. Multi-level optimization of building design, energy system sizing and operation. Energy 90, 1775–1789. https://doi.org/10.1016/j.energy.2015.07.007

Feurer, M., Hutter, F., 2019. Hyperparameter optimization, in: Automated Machine Learning. Springer, Cham, pp. 3–33.

Geoffrion, A.M., 1972. Generalized Benders decomposition. J. Optim. Theory Appl. https://doi.org/10.1007/BF00934810

Grimstad, B., Andersson, H., 2019. ReLU networks as surrogate models in mixed-integer linear programs. arXiv.

Grossmann, I.E., 2013. Global optimization in engineering design. Springer Science & Business Media.

Gupta, T.K., Raza, K., 2019. Optimization of ANN Architecture: A Review on Nature-Inspired Techniques, in: Dey, N., Borra, S., Ashour, A.S., Shi, F. (Eds.), Machine Learning in Bio-Signal Analysis and Diagnostic Imaging. Academic Press, pp. 159–182. https://doi.org/10.1016/b978-0-12-816086-2.00007-2

Haeffele, B.D., Vidal, R., 2017. Global optimality in neural network training, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7331–7339.

Hassibi, B., Stork, D., 1993. Second order derivaties for network prunning: Optimal brain surgeon. Adv. NIPS5 164–171.

He, Y., Dong, X., Kang, G., Fu, Y., Yan, C., Yang, Y., 2019. Asymptotic Soft Filter Pruning for Deep Convolutional Neural Networks. IEEE Trans. Cybern. 1–11. https://doi.org/10.1109/tcyb.2019.2933477

Hossain, K.M.A., 2014. Predictive Ability of Improved Neural Network Models to Simulate Pollutant Dispersion. Int. J. Atmos. Sci. 2014, 1–12. https://doi.org/10.1155/2014/141923

Huang, H., An, H., Ma, H., Chen, S., 2019. An engineering method for complex structural optimization involving both size and topology design variables. Int. J. Numer. Methods Eng. 117, 291–315. https://doi.org/10.1002/nme.5957

Jensen, C.A., El-Sharkawi, M.A., Marks, R.J., 2001. Power system security assessment using neural networks: Feature selection using fisher discrimination. IEEE Trans. Power Syst. 16, 757–763. https://doi.org/10.1109/59.962423

Jerez, I.J., Muñoz, F., Gomez, J.M., 2014. Approach to a reliable solution strategy for performing phase equilibrium calculations using MINLP optimization, Latin American Applied Research.

Joy, P., Djelassi, H., Mhamdi, A., Mitsos, A., 2019. Optimization-based global structural identifiability. Comput. Chem. Eng. 128, 417–420.

Kalyan, K., Jakhia, B., Lele, R.D., Joshi, M., Chowdhary, A., 2014. Artificial Neural Network Application in the Diagnosis of Disease Conditions with Liver Ultrasound Images. Adv. Bioinformatics 2014. https://doi.org/10.1155/2014/708279

Kavzoglu, T., Mather, P.M., 2002. The role of feature selection in artificial neural network applications. Int. J. Remote Sens. 23, 2919–2937. https://doi.org/10.1080/01431160110107743

Kavzoglu, T., Mather, P.M., 1999. Pruning artificial neural networks: An example using land cover classification of multi-sensor images. Int. J. Remote Sens. 20, 2761–2785. https://doi.org/10.1080/014311699211796

Kavzoglu, T., Mather, P.M., 1998. Assessing artificial neural network pruning algorithms, in: Proceedings of the 24th Annual Conference and Exhibition of the Remote Sensing Society. Greenwich, UK, pp. 603–609.

Kesavan, P., Allgor, R.J., Gatzke, E.P., Barton, P.I., 2004. Outer approximation algorithms for separable nonconvex mixed-integer nonlinear programs. Math. Program. 100, 517–535.

Kim, K.J., Han, I., 2000. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Syst. Appl. 19, 125–132. https://doi.org/10.1016/S0957-4174(00)00027-0

Kocis, G.R., Grossmann, I.E., 1988. Global optimization of nonconvex mixed-integer nonlinear programming (MINLP) problems in process synthesis. Ind. Eng. Chem. Res. 27, 1407–1421.

Kronqvist, J., Bernal, D.E., Lundell, A., Grossmann, I.E., 2019. A review and comparison of solvers for convex MINLP. Optim. Eng. 20, 397–455.

Lawler, E.L., Wood, D.E., 1966. Branch-and-Bound Methods: A Survey. Oper. Res. 14, 699–719. https://doi.org/10.1287/opre.14.4.699

Le, C., 1990. Optimal brain damage, in: Advances in Neural Information Processing Systems. Denveri Colorado, USA, pp. 598–605.

Lee, J.H., Shin, J., Realff, M.J., 2018. Machine learning: Overview of the recent progresses and implications for the process systems engineering field. Comput. Chem. Eng. 114, 111–121. https://doi.org/10.1016/j.compchemeng.2017.10.008

Levasseur, L.P., Hezaveh, Y.D., Wechsler, R.H., 2017. Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing. arXiv 850, L7. https://doi.org/10.3847/2041-8213/aa9704

Lodi, A., Nagarajan, V., 2019. Integer Programming and Combinatorial Optimization: 20th International Conference, IPCO 2019, Ann Arbor, MI, USA, May 22-24, 2019, Proceedings. Springer.

Manngård, M., Kronqvist, J., Böling, J.M., 2018. Structural learning in artificial neural networks using sparse optimization. Neurocomputing 272, 660–667.

McCulloch, W.S., Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys. 5, 115–133.

Miao, J., Niu, L., 2016. A Survey on Feature Selection. Procedia Comput. Sci. 91, 919–926. https://doi.org/10.1016/j.procs.2016.07.111

Misener, R., Floudas, C.A., 2014. ANTIGONE: algorithms for continuous/integer global optimization of nonlinear equations. J. Glob. Optim. 59, 503–526.

Okuyucu, H., Kurt, A., Arcaklioglu, E., 2007. Artificial neural network application to the friction stir welding of aluminum plates. Mater. Des. 28, 78–84. https://doi.org/10.1016/j.matdes.2005.06.003

Pintarič, Z.N., Kravanja, Z., 2000. The two-level strategy for MINLP synthesis of process flowsheets under uncertainty, in: Computers and Chemical Engineering. pp. 195–201. https://doi.org/10.1016/S0098-1354(00)00512-3

Poernomo, A., Kang, D.-K., 2018. Biased Dropout and Crossmap Dropout: Learning towards effective Dropout regularization in convolutional neural network. Neural Networks 104, 60–67. https://doi.org/https://doi.org/10.1016/j.neunet.2018.03.016

Rössig, A., Petkovic, M., 2021. Advances in verification of ReLU neural networks. J. Glob. Optim. 81, 109–152.

Sahinidis, N. V., 1996. BARON: A general purpose global optimization software package. J. Glob. Optim. 8, 201–205. https://doi.org/10.1007/bf00138693

Schlueter, M., 2014. MIDACO software performance on interplanetary trajectory benchmarks. Adv. Sp. Res. 54, 744–754. https://doi.org/10.1016/j.asr.2014.05.002

Schweidtmann, A.M., Mitsos, A., 2019. Deterministic Global Optimization with Artificial Neural Networks Embedded. J. Optim. Theory Appl. 180, 925–948. https://doi.org/10.1007/s10957-018-1396-0

Sietsma, J., Dow, R.J.F., 1988. Neural net pruning - why and how, in: IEEE International Conference on Neural Networks. IEEE, pp. 325–333. https://doi.org/10.1109/icnn.1988.23864

Smith, E.M.B., Pantelides, C.C., 1999. A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. Comput. Chem. Eng. 23, 457–478.

Stamoulis, D., Cai, E., Juan, D.-C., Marculescu, D., 2018. Hyperpower: Power-and memory-constrained hyper-parameter optimization for neural networks, in: 2018 Design, Automation & Test in Europe Conference & Exhibition (DATE). IEEE, pp. 19–24.

Venkatasubramanian, V., 2019. The promise of artificial intelligence in chemical engineering: Is it here, finally? AIChE J. 65, 466–478. https://doi.org/10.1002/aic.16489

Verikas, A., Bacauskiene, M., 2002. Feature selection with neural networks. Pattern Recognit. Lett. 23, 1323–1335. https://doi.org/10.1016/S0167-8655(02)00081-8

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. 106, 25–57.

Wen, C., Ma, X., 2008. A max-piecewise-linear neural network for function approximation. Neurocomputing 71, 843–852. https://doi.org/10.1016/j.neucom.2007.03.001

Xie, X., Zhang, H., Wang, Junze, Chang, Q., Wang, Jian, Pal, N.R., 2019. Learning optimized structure of neural networks by hidden node pruning with L1 regularization. IEEE Trans. Cybern. 50, 1333–1346.

Zhou, Y., Yen, G.G., Yi, Z., 2019. A Knee-Guided Evolutionary Algorithm for Compressing Deep Neural Networks. IEEE Trans. Cybern. 1–13. https://doi.org/10.1109/tcyb.2019.2928174

2
Data-driven Modeling of an Industrial Ethylene Oxide Plant: Superstructure-based Optimal Design for Artificial Neural Networks

Hasan Sildir,[a*] Sahin Sarrafi,[b] Erdal Aydin[c]

[a]*Department of Chemical Engineering, Gebze Technical University, Kocaeli 41400, Türkiye*
[b]*SOCAR Türkiye, R&D and Innovation Co., Aliaga, Izmir, 35800, Türkiye*
[c]*Department of Chemical Engineering, Bogazici University, Bebek, İstanbul 34342, Türkiye*
hasansildir@gtu.edu.tr

Conferance: The 31st European Symposium on Computer Aided Process Engineering (ESCAPE 31), Istanbul, Türkiye

Journal: Computer Aided Chemical Engineering, Elsevier, Volume 50, **2021**, 445-450

https://www.sciencedirect.com/science/article/abs/pii/B978032388506550070X

DOI: https://doi.org/10.1016/B978-0-323-88506-5.50070-X

Abstract

Optimum selection of input variables, number of hidden neurons and connections between the network elements delivers the best configuration of an artificial neural network (ANN), resulting in reduced over-fitting and improved performance. In this study, a superstructure-oriented ANN design and training algorithm is suggested and implemented on an industrial Ethylene Oxide (EO) plant for the prediction of product related variables (i.e. EO production rate). Proposed formulation is a mixed integer nonlinear programming problem (MINLP), which takes the existence of inputs, neurons and connections of the network into account by binary variables in addition to continuous weights of existing connections. Investigations show that almost 90% of the connections are removed compared to the fully connected ANN (FC-ANN) with 50% decrease in the number of inputs of the ANN, approximately. The modified ANN delivers a better prediction performance over FC-ANN, which suffers from over-fitting.

Keywords: machine learning; artificial neural networks; superstructure optimization; process modeling; mixed integer nonlinear programming

Introduction

Ethylene oxide (EO), a captive product used for the production of ethylene glycols (MEG, DEG, and TEG), is produced through selective oxidation of ethylene and oxygen in the presence of an Ag supported α-Al2O3 catalyst in the EO/EG plant located in the SOCAR Türkiye/Petkim Petrochemical plant

in Türkiye. The EO process takes place in two parallel fixed-bed multi-tubular reactors at a temperature and pressure range of 240–260 °C and 17–18 bar, respectively. Both ethylene epoxidation and EO combustion occur in the reactors and only the former is desirable. Side reactions reduce the EO yield and an increase in CO_2 emission from the plant. From both safety and optimum control perspectives, the monitoring of the concentrations of the streams, which are measured through two separate online gas chromatography (GC) instruments, is of great importance. The former is designed for hydrocarbons and permanent gases whereas the latter focuses on the chlorine-based compounds. The operating strategy of the reactors is to maximize the EO yield under various operational constraints. The simplified process flow diagram of the process is shown in Fig. 1.

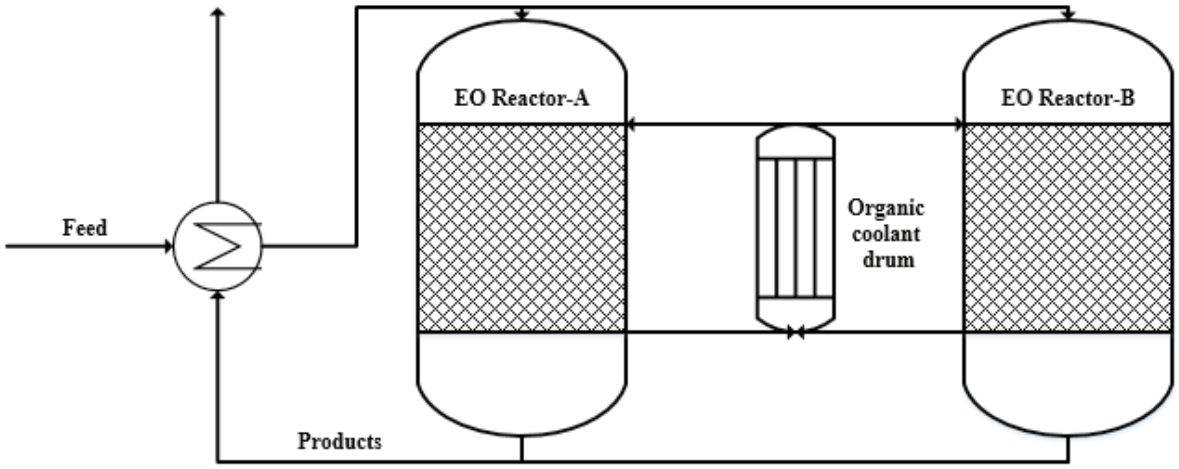

Figure 1. Simplified process flow diagram of EO process

High number of criteria is considered during EO production at the desired levels under the continuous and sluggish catalyst deactivation. These criteria lay the foundation of the features of information matrix exploited throughout this study as detailed in Table.1. All the feed compositions are measured before the reactor (i.e., after EO and CO_2 absorption section). Fresh VCM feed is introduced into the reactor inlet stream right before the reactor. Real-time optimization and automation of such a complex process is a challenging task, requiring significant effort in first principles and mechanistic modeling. Alternatively, Artificial Neural Networks (ANNs) are promising empirical models to estimate the product related variables from easily measurable variables (i.e. temperatures and pressures).

Table 1. Descriptions of the features and inputs

TAG Description	Unit	TAG Description	Unit
Lean absorbent flow rate	t/h	Fresh C_2H_4 feed	t/h
CO_2 absorber overhead pressure	kg/ cm^2g	Fresh CH_4 feed	kg/h
Feed composition (C_2H_4)	% (v/v)	CH_3Cl (methyl chloride) in the recycle	ppm
Feed composition (O_2)	% (v/v)	VCM (Viniyl chloride monomer) in the recycle	ppm
Feed composition (N_2)	% (v/v)	C_2H_5Cl (ethyl chloride) in the recycle	ppm
Feed composition (Ar)	% (v/v)	Reactor A coolant pressure	kg/cm^2g
Feed composition (CH_4)	% (v/v)	Reactor A inlet pressure	kg/cm^2g
Feed composition (C_2H_6)	% (v/v)	Flow rate to reactor A	t/h
Feed composition (CO_2)	% (v/v)	Fresh VCM feed to reactor A	kg/h
Feed composition (EO)	% (v/v)	Product composition (EO)	% (v/v)
Fresh O_2 feed	t/h		

Methodology: Optimal Superstructure-based Design of ANNs

The standard application of ANNs includes using fully connected networks, where all inputs, neurons and outputs are entirely connected. Fully connected ANN architectures (FC-ANNs) have high number of parameters. Typically, as the dimensions get larger, higher number of connections and parameters are introduced. It is usually expected that the increase in the number of parameters provides higher capability of fitting to the training data. Nevertheless, it is worthwhile here to mention that this increase may easily result in overfitting, and therefore poor prediction capability of the ANN model. Introducing more data is usually not a satisfactory effort as new data do not carry additional statistical information unless they are collected from a different location in the plant. A feedforward ANN model is generally expressed as:

$$y = f_1(A \cdot f_2(B \cdot u + C) + D) \quad (1)$$

where f_1 and f_2 are output and hidden layer activation functions respectively. A typical activation function is the hyperbolic tangent function, which is used in this study as well. A and B are weight matrices; C and D are bias vectors; u is the input vector and y is the output vector. Related ANN parameters are all continuous, and theoretically unbounded, and their dimensions depend on the number of inputs, outputs, and number of neurons which is determined manually before training. Identification issues of these ANN parameters may result in poor testing performance. A typical solution to this problem is to include a regularization term to penalize the large values of ANN parameters,

which unfortunately cannot regularize the hyper parameters ([1]). Pruning is another method to reduce the number of connections in ANNs ([2-3]). Dua proposed using a general mixed-integer optimization formulation to eliminate overfitting by detecting the optimal configuration of ANNs ([4]). Both number of neurons and existence of the interconnections are included in the objective function. Yet, this formulation does not consider selection of optimum input variables. Moreover, resulting formulations are either mixed-integer linear programs with fixed parameter weights or small scale MINLPs with fixed structures, number of nodes and/or interconnections, exhibiting poor flexibility. On the other hand, this significant study shows that much better test, and thus prediction performance can be achieved with fewer neurons and connections.

In this study, a novel MINLP (mixed-integer nonlinear programming) formulation is developed for the design and training of an optimal architecture feedforward ANN (OA-ANN), by modifying traditional ANN equations. The MINLP formulation introduces additional binary variables to the traditional ANN equations to represent the existence of network elements. The resulting superstructure formulation also takes the selection of the input variables into account in addition to the number of hidden neurons and connections, which in turn maximizes the overall flexibility and strength of the formulation. Accordingly, the proposed MINLP problem, whose objective is to minimize the number of connections of the ANN, is given by:

$$\min_{A,A_{binary},B,B_{binary},C,D,N_{binary},U_{binary}} \sum_{i,j} A_{binary,ij} + B_{binary,ij}$$

$$s.t.\quad F = \sum_{i=1}^{N} \| f_1((A \circ A_{binary}) \cdot diag(N_{binary}) \cdot f_2((B \circ B_{binary}) \cdot diag(U_{binary}) \cdot u_i + C) + D) - y_i \|$$

$$F \leq \mathcal{E} \tag{2}$$

$$A_{binary,ij} \leq N_{binary,j}$$

$$B_{binary,ij} \leq U_{binary,j}$$

$$A_{i,j}, B_{i,j}, C, D \in [-4,4]$$

$$A_{binary}, B_{binary}, N_{binary}, U_{binary} \in \{0,1\}$$

where $\circ$ is the Hadamard product operator; u_i and y_i are the i^{th} input and output sample respectively. N is the number of samples used for the training. $A_{binary,ij}$ and $B_{binary,ij}$ are matrices with binary values representing the existence of connections. N_{binary} and U_{binary} are the binary vectors defining the existence of a neuron and an input, respectively. The existence of a particular connection between a neuron and an input is defined by the binary variable $B_{binary,ij}$. A_{ij} is the continuous weight parameter of the connection between the j^{th} neuron and the i^{th} output. Similarly, B_{ij} represents the connection between an input and the corresponding neuron. $\mathcal{E}$ is the upper bound for the overall training error. This way, a trade-off is included into the training, which in turn is expected to increase the test performance after implementing the proposed formulation. Please note that suggested formulation has the flexibility to result in a fully

connected network once ε is set to a small value. In other words, a fully connected ANN is the upper bound of the suggested formulation. Problem (2) is a non-convex, and generally a large-scale MINLP, which is quite challenging to solve to the global optimum. In this work, an adaptive and evolutionary algorithm is used to solve the resulting non-convex MINLP problem. This method has a similar idea with [5] and it decomposes the original MINLP into an integer program (IP) and a nonlinear program (NLP). IP's only include binary decision variables that can be adjusted during optimization, whereas NLP's only involve continuous decision variables. The IP stands on the outer loop and is solved via the genetic algorithm based IP solver of Matlab while the inner loop NLP is solved by using IPOPT ([6]). Two problems are solved sequentially until the tolerance value of the original problem objective value or the maximum wall clock time is reached.

Please note that IPOPT is a local solver and might add randomness to the outer problem as the IP solution is based on a black-box genetic algorithm. Thus, we do not propose any integer cuts to be provided during the iterations. In addition to that, sensitivity analysis of the proposed method might be of interest to fully assess the regularization of the optimal ANNs using heuristic solutions, which is left as a future study. A global NLP solver can also be used for the inner problem to deal with the aforementioned challenges. On the other hand, investigations show that suggested method usually provides a sufficient heuristic solution in less than 3 min.

Results and Discussion

The proposed MINLP based approach is tested on the industrial EO plant. Corresponding data cover hourly measurements of 105 randomly selected days. We tested different training ratios, e.g. 20%, 50%, and 70%, to demonstrate the impact of the approach. OA-ANN architecture from the training with 20% of the data is shown in Fig. 2 as a typical demonstration. In the design of OA-ANN with 20% of the data for training, 9 inputs, out of 20, are selected from which connections exist to hidden layer as shown in Fig. 2. The thickness of the connections represents the absolute magnitude of the corresponding weight. Note that only significant connections are maintained, eliminating the ineffective connections and thereby tightening the non-linear training optimization problem.

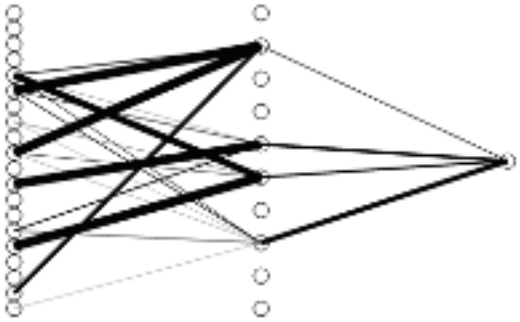

Figure 2. OA-ANN for 20% training data

In addition, the OA-ANN architecture, with only 4 hidden layer neurons, does not have a fully connected architecture as the selected inputs do not have connection to all hidden neurons. As a result, a more efficient information flow is obtained through the network. On the other hand, there are many different architectures which may deliver a similar performance since the MINLP problem is highly non-convex and inherently contains input multiplicity both due to input variable correlations and parameter dependencies. Accordingly, the heuristic solution method and even the optimization algorithm tuning parameters might result in different architectures. The selected inputs include the lean absorbent flow rate, absorber overhead pressure, feed composition of N_2 and Ar, feed CO_2 composition, fresh O_2 feed, fresh CH_4 feed, Vinyl chlorine monomer and ethyl chloride feed. Note that some input variables show significant correlation and there are many subset combinations to deliver a similar performance. Thus, the selected inputs, likewise the structure in Fig. 2, is not a unique solution; different initial guesses and MINLP solution approaches would result in a different optimal subset. Fig. 2 also reveals the impact of the input variables on the network, which is represented by the thickness of the connection and scaled by the corresponding weight. For instance, N_2 composition in the feed contributes to the network calculations in smaller scale compared to the chlorine compounds, which have a significant impact on the process control. A similar observation is also valid for O_2 feed, as well.

Detailed training and test performance comparison of FC-ANN and OA-ANN at various training ratios are presented in Table 2 which includes some common statistical measures.

Table 2. OA-ANN and FC-ANN detailed performance comparison

Train ratio	20%				50%				70%			
	OA-ANN		FC-ANN		OA-ANN		FC-ANN		OA-ANN		FC-ANN	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
RMSE	0.003	0.004	0.0001	0.0140	0.003	0.0035	0.0004	0.031	0.0045	0.005	0.0004	0.047
CV	0.046	0.061	0.008	0.147	0.051	0.061	0.018	0.25	0.065	0.093	0.017	0.296
MAE	-	0.424	-	2.481	-	0.49	-	4.53	-	0.37	-	1.03

Table 2 includes root mean square error (RMSE), coefficient of variation (CV) and maximum absolute error (MAE) as a comparison metric. Due to the eliminated over-fitting, the training and the test performances of OA-ANN show a good agreement at all training ratios. Although FC-ANN demonstrates significantly lower training error with high number of connections, it suffers from larger test error compared to OA-ANN. Similar insight can also be obtained from CVs. In all cases, test performance of OA-ANN is superior to FC-ANN despite higher training error. The architecture comparisons are given in Table 3. The network dimensions of OA-ANN are significantly less than FC-ANN in all training ratios, providing additional computational advantages when ANN update is necessary. Overfitting causes poor test performance as shown in Fig. 3. This issue stemms from high number of connections in FC-ANN, whose weights could not be identified based on the statistical content of the current training data. Both FC-ANN and OA-ANN test results are demonstrated in addition to normalized measurements in Fig. 3. Note that the OA-ANN has a better capability to predict a different operating condition, as shown between sample numbers 400-500.

	OA- ANN			FC-ANN
	20%	50%	70%	-
Neurons	4	3	5	10
Connections	22	19	26	210
Inputs	9	9	8	20

Table 3. OA-ANN and FC-ANN architecture comparison

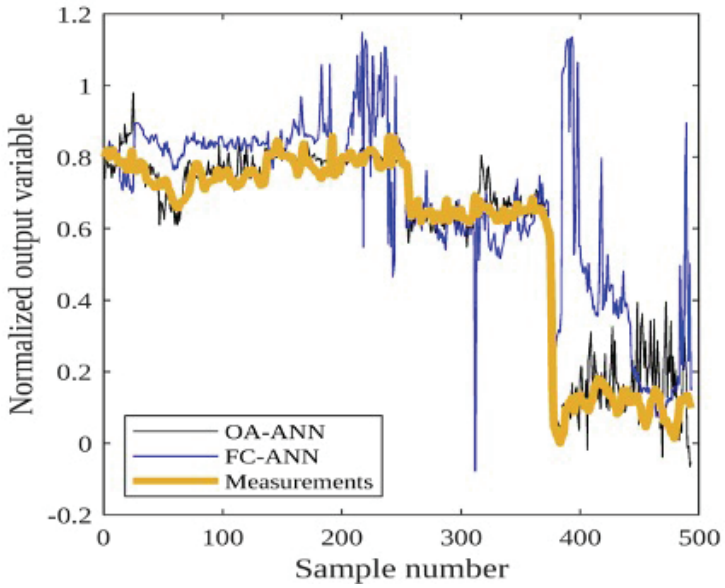

Figure 3. Test data performance of OA-ANN and FC-ANN

Conclusion

The design and synthesis of ANNs is a challenging and important task. Ineffectively large networks provide significant test errors due to over-fitting. Test performance of the proposed approach is significantly better compared to the FC-ANN, despite using approximately 10% of the connections and half of the inputs only. Reduced input space is advantageous for the real-time operation when model update or sensor failures are under consideration. Furthermore, the OA-ANN is superior in various training ratios, more robust to extrapolation and prediction of different operating conditions. The heuristic MINLP solution algorithm provides a satisfactory network architecture and still has theoretical potential for development. Our current focus includes the implementation to actual plant.

Acknowledgements

This publication has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C245). However, the entire responsibility of the publication belongs to the owner of the publication.

References

[1] M. Manngård, J. Kronqvist, and J. M. Böling, "Structural learning in artificial neural networks using sparse optimization," Neurocomputing, vol. 272, pp. 660–667, 2018.

[2] Y. Zhou, G. G. Yen, and Z. Yi, "A Knee-Guided Evolutionary Algorithm for Compressing Deep Neural Networks," IEEE Trans. Cybern., pp. 1–13, 2019.

[3] Xie, H. Zhang, J. Wang, Q. Chang, J. Wang, and N. R. Pal, "Learning Optimized Structure of Neural Networks by Hidden Node Pruning with L1 Regularization," IEEE Trans. Cybern., vol.50, no. 3, pp. 1333–1346, 2020.

[4] V. Dua, "A mixed-integer programming approach for optimal configuration of artificial neural networks," Chem.Eng. Res. Des., vol. 88, no. 1, pp. 55–60, 2010.

[5] Pintaric, Z.N.; Kravanja, Z. The two-level strategy for MINLP synthesis of process flowsheets under uncertainty. Comput. Chem. Eng. 2000, 24, 195–201.

[6] Waechter, A., and L. T. Biegler. On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming, volume 106. 2006.

3
Uncertainty Propagation based MINLP approach for Artificial Neural Network Structure Reduction

Hasan Sildir[a,*], Sahin Sarrafi[b] and Erdal Aydin[c, d, e]

[a] *Department of Chemical Engineering, Gebze Technical University, Gebze, 41400 Türkiye,*
[b] *SOCAR Türkiye R&D and Innovation Co., Aliaga, Izmir, 35800 Türkiye*
[c] *Department of Chemical and Biological Engineering, Koc University, Sarıyer, Istanbul 34450 Türkiye.*
[d] *Koc University TUPRAS Energy Center, Rumelifeneri, Sarıyer, Istanbul 34450 Türkiye*
[e] *Department of Chemical Engineering, Bogazici University, Bebek, Istanbul 34342 Türkiye*
hasansildir@gtu.edu.tr

Journal: Processes **2022**, 10(9), 1716;

https://www.mdpi.com/2227-9717/10/9/1716

DOI: https://doi.org/10.3390/pr10091716

Abstract

The performance of artificial neural networks (ANNs) is highly influenced by the selection of input variables and the architecture defined by hyper parameters such as the number of neurons in the hidden layer and connections between network variables. Although there are some black-box and trial and error based studies in the literature to deal with these issues, it is fair to state that a rigorous and systematic method providing global and unique solution is still missing. Accordingly, in this study, a mixed integer nonlinear programming (MINLP) formulation is proposed to detect the best features and connections among the neural network elements while propagating parameter and output uncertainties for regression problems. The objective of the formulation is to minimize the covariance of the estimated parameters while by i) detecting the ideal number of neurons, ii) synthesizing the connection configuration between those neurons, inputs and outputs, and iii) selecting optimum input variables in a multi variable data set to design and ensure identifiable ANN architectures. As a result, suggested approach provides a robust and optimal ANN architecture with tighter prediction bounds obtained from propagation of parameter uncertainty, and higher prediction accuracy compared to the traditional fully connected approach and other benchmarks. Furthermore, such a performance is obtained after elimination of approximately 85% and 90% of the connections, for two case studies respectively, compared to traditional ANN in addition to significant reduction in the input subset.

Keywords: Artificial Neural Networks; Error Propagation; Mixed Integer Nonlinear Programming; Optimal Input Selection; Parameter Uncertainty

Introduction

Energy producers are globally struggling with the inadequacy of energy resources and energy efficiency problems, which have become even more active in recent years due to COVID-19 pandemics. Another issue to be addressed in today's world is energy efficiency which is a significant economic growth and prosperity metric. In particular, optimal energy production, distribution, and consumption have vital importance on the economic independence due to the increased energy efficiency. Accordingly, artificial intelligence (AI) and machine learning (ML) have emerged to be nice candidates to improve the efficiency of energy systems without requiring mass investments as opposed to converting all of the energy processing units. Typically, a software can be linked to an energy system and can alter its performance, and thus energy efficiency, using the state-of-the-art approaches on AI and ML.

Artificial Neural Networks (ANNs) are sophisticated models being able to represent complex relationships between inputs and outputs of a certain system/process. ANNs are data driven and, in particular cases, considered as an alternative to first principles-models because it is challenging to derive and validate such models due to unknown driving forces in the process and lack of spatial measurements [1]. The foundations of ANNs are laid by Mcculloch and Pitts in [2], in which the neural activity working mechanisms are discussed. Significant theoretical advancements have been achieved and these achievements have resulted in a wide range of applications showing promising performance [3–5].

Various research activities have also led to different terminologies and applications such as artificial intelligence, machine learning and deep learning [6–9]. Murugan and Natarajan designed a dynamic soft sensor to estimate the biomass concentration in a complex pilot plant from easily measurable plant variables (i.e. pH) [10]. Kaur and Kumari used ANNs to detect patterns and risks for diabetes [11]. The applications in mineral processing are discussed in [12]. Moreover, ANN related methods have found applications in energy economics and finance [13]. In [14], an ANN model is developed to forecast carbon emissions from several macroeconomic indicators such as economic growth. Air pollution forecast studies have used ANNs widely as well [15]. ANNs are also employed in advanced process automation technologies such as stochastic model predictive control [16]. MacMurray and Himmelblau showed the importance of nonlinear and complex processing capability of ANNs in a packed distillation column [17]. Biswas *et al.* modeled the energy consumption of residential sector through ANNs and showed a good match between prediction and measurements [18]. In addition, there are several successful applications in the literature in terms of using meta-heuristic and multi-verse optimization methods for the improvement of machine learning

models in general, resulting in more stable models. Such meta-heuristic based hyper-parameter tuning methods might also improve the capacity of neural networks if applied [19,20]. Ardiansyah et al. used ANNs for the prediction of quality variables and design strategy for an extraction process [21].

Traditional fully connected ANN architectures (FC-ANNs), which are defined by hyper parameters, are composed of single hidden layer in addition to input and output layers. Usually, a trial and error procedure on hyper-parameters is applied until a satisfactory training performance is obtained. Also, the number of connections, and the number of tuning parameters increase with the number of inputs, outputs and neurons. As a result, FC-ANNs suffer from parameter identifiability issues due to multiple solutions, lack of accountable measurements and over parametrization [22]. Overfitting and parameter identifiability problems result in large prediction bounds and therefore poor prediction performance, reducing the performance significantly especially in complex systems. Note that there are some alternative methods to eliminate some of the aforementioned issues using statistical measures [23]. The resulting uncertainty is also addressed in [24] using bayesian computation.

Group Method of Data Handling (GMDH) [25] develops relatively smaller polynomial models for the approximation of more complex models through elimination of unrelated variables. At the same time, GMDH can be classified as a more sophisticated pruning method. An external selection criterion, which is a nontrivial task to formulate with many different alternatives [26], is used to define the existence or the elimination of a particular network variable. Then, a sorting procedure [27] is applied for the selection of the best architecture among many generated networks. Due to sorting, GMDH can be classified under sequential approaches focusing on the elimination of neurons and inputs. In addition, GMDH does not include covariance of parameters into the objective function.

Dua suggested solving a general mixed-integer optimization problem to eliminate the connections of ANNs during training [28]. Both number of neurons and existence of the interconnections are included in the objective function to be minimized together with the training error. However, this formulation does not consider the parameter uncertainty and the selection of optimum input variables. Resulting problem formulations in the case studies are either mixed-integer linear programs with fixed parameter weights -which are significantly easier problems to solve- or small scale MINLPs with fixed structures, fixed number of nodes and/or interconnections of the ANNs. Commercially available programming language GAMS (General Algebraic Modeling System) is used to solve the corresponding optimization problems. This significant contribution shows that increased performance can be achieved with fewer number of neurons and connections in ANNs. Similarly, in recent studies, the authors

showed that optimal structure detection for ANNs can also be realized using more flexible mixed integer nonlinear programming and piece-wise linear formulations [29–32].On the other hand, it is vital to state here that none of these formulations include the parameter uncertainty covariance, which is again a critical measure for reliable training and reduced overfitting for not only ANNs but also for almost all types of machine learning applications for regression. Thus, for truly effective structure detection for ANNs, uncertainty effect must also be incorporated into the existing MINLP formulations proposed by the authors.

Another significant aspect of ANN training is the selection of optimal input variables from a complex data set [33]. Usually, the leading signals are not known and the data set contains correlated or redundant variables. In such cases, the optimal selection of input variables becomes an important issue [34], which in turn calls for a robust method to reduce the number of ANN parameters and the input subset selection to provide a more robust identification.

Sun et. al utilized genetic algorithm for automatic design of Convolutional Neural Network architectures for image classification [35]. Benardos and Vosniakos proposed a genetic algorithm to modify the ANN architecture [36]. In a similar fashion, Dua developed a general mixed-integer program to eliminate some connections of ANNs during training [28]. Both the number of neurons and the existence of the interconnections are included in the multi objective function in addition to training error. Yet, these formulations do not consider the parameter uncertainty as a measure of overfitting and the reduction in the input space. On the other hand, it has been shown that a similar performance can be obtained with fewer neurons and connections.

In this study, a novel MINLP (mixed-integer nonlinear programming) formulation is developed for the automatic synthesis and training of an optimal feedforward ANN architecture (OA-ANN). Traditional ANN equations are modified and the training procedure considers the parameter uncertainty to eliminate overfitting. Main contributions of the proposed work are: i) detecting ideal number of neurons and selection of inputs by introducing binary variables in the MINLP formulation for regression problems through a heuristic yet tailored solution algorithm and ii) synthesizing the optimal information flow between neurons, inputs and outputs are characterized by introducing binary variable matrices as A_{binary} and B_{binary} while minimizing the overfitting criterion by minimizing the parameter covariance as another objective in the optimization for regression problems to account for the tightening the prediction bounds of continuous output variables. To the best of authors' knowledge such an approach does not exist in the literature. Moreover, to show the potential of improvement for energy systems, a case study about a strong data set on energy consumption predictions is considered in this work.

The paper is structured as follows: In II, the derivation of the ANN (OA-ANN) expressions and the solution algorithm for the corresponding MINLP program is explained. Results of the proposed formulation and comparisons to FC-ANN and literature benchmarks are provided in III. Finally, section IV concludes this study.

Materials & Methods

A typical feedforward ANN expression is given by:

$$y = f_1(A \cdot f_2(B \cdot u + C) + D) \tag{1}$$

where f_1 and f_2 are output and hidden layer activation functions respectively; A and B are weight matrices; C and D are bias vectors; u is input vector and y is output vector. Note that identity activation function is used in this formulation at the input layer and it is not shown in (1) explicitly. The continuous ANN parameters A, B, C and D are estimated from preferably high number of samples. The dimensions of those parameters depend on the number of inputs, outputs and number of neurons (a hyper parameter), which is determined manually before training. In general, as the dimensions get larger, higher number of connections and parameters are introduced, which in turn provides higher capability of fitting to the training data.

The architecture given by (1) represents a FC-ANN which transforms the information in input, u, to the succeeding layers, and eventually to the output vector, y. Addition of a higher number of hidden layers is a straightforward mathematical task as more parameters, connections and neurons are introduced. This task is in the concept of deep learning, providing useful results in the literature [9,37].

FC-ANNs are traditionally trained through nonlinear optimization using the following objective function:

$$Min_{A,B,C,D} \sum_{i=1}^{N} \| f_1(A \cdot f_2(B \cdot u_i + C) + D) - y_i \| \tag{2}$$

where u_i and y_i are the i^{th} input and output sample respectively; and N is the number of samples used for the training.

Equation (2) takes the training error into account only and does not consider the parameter identifiability or architecture efficiency issues. However, the practical and structural limitations on the estimation of those parameters are vital in order to increase overall model quality and prediction robustness, and

to reduce overfitting. Otherwise, some parameters might have little impact on the output while exhibiting strong correlations among other parameters, making it almost impossible to identify them uniquely [38] despite significant computational load. The outcome of such problem would be the large variance in the ANN predictions due to the parameter uncertainty propagation to outputs and the poor prediction accuracy in the test data once there is a significant difference between training and the test data. In addition, the parameter correlation, which is caused by inefficient model architecture and high number of parameters, results in significant computational load during training or model update as optimization algorithm calculates similar objective function with distinct decision variable values although new data are collected for model correction in real time. An alternative straightforward method to avoid the aforementioned problems would be to include more training data, but is practically not useful mostly once the data are not measured in a distinct data regime. Another alternative is the modification of the model itself either by lumping some parameters, removing some of them by a statistical measure or fixing some of them to a particular value to reduce the parameter correlation, thereby eliminate overfitting. However, this method cannot be considered as automatic and requires significant manual effort.

One of the significant contributions of this work is the integration of the parameter uncertainty propagation together with the proposed MINLP method, which will be discussed later. Parameter covariance matrix is a measure of identifiability in complex models. Based on the Cramer and Rao theorem [39], the inverse of the Fisher Information Matrix (*FIM*) is a lower bound for parameter covariance matrix:

$$cov(\hat{p} - p^{actual}) \geq FIM^{-1} \tag{3}$$

where $\hat{p}$ is the vector of estimated parameters; p^{actual} the actual value of the parameters; *FIM* is calculated from:

$$FIM = \frac{1}{\sigma^2} JJ^T \tag{4}$$

where σ^2 is the variance of the output error; J is the parameter sensitivity matrix which is evaluated at a particular point. Small eigenvalues of *FIM* deliver large lower bounds for the parameters, which theoretically means that all parameters cannot be identified uniquely.

The parameter uncertainty can be propagated to the outputs through the traditional error propagation formulation [40]:

$$cov_y = J \cdot cov_p \cdot J^T \tag{5}$$

where cov_y is the covariance matrix of outputs; cov_p is the covariance matrix of parameters. Diagonal values of cov_y and cov_p provides an intuitive understanding of the uncertainty since each element is the variance of the corresponding variable. From ANN perspective, once the values in cov_p decrease, corresponding ANN predictions deliver a tighter uncertainty range, resulting in a more robust and reliable prediction generally [41]. As a result, the selected features and connections of the neural network will provide more robust prediction capability.

The modified ANN (OA-ANN) equation to be taken into account as opposed to the standard formulation is given as:

$$y = f_1\left(\left(A \circ A_{binary}\right) \cdot diag(P) \cdot f_2\left(\left(B \circ B_{binary}\right) \cdot diag(U) \cdot u + C\right) + D\right) \quad (6)$$

where $\circ$ is the Hadamard product (element-wise multiplication) operator; A_{binary} and B_{binary} are matrices with binary elements, representing the connection existence of hidden layer neurons with the output layer and hidden layer with the input layer respectively; P is a binary vector which represents the existence of neurons; U is a binary vector which represents the input selection; f_1 and f_2 are hyperbolic tangent activation function in this paper. Furthermore, please note that f_1 and f_2 are usually decided before training manually. Even though we propose to use hyperbolic activation functions in this study, suggested framework is also extendable to take into account the type of the activation functions as decision variables.

Mixed-integer programming typically considers the continuous and discrete decisions together to implement an optimization objective subject to constraints. For neural networks, the existence (or non-existence) off the features must be represented as a discrete, binary (0-1) decision variable whereas the corresponding weight values for training are continuous. The training formulation of OA-ANN is an MINLP problem and is given by:

$$
\begin{gathered}
Min_{A,A_{binary},B,B_{binary},C,D,P,U} \sum diag\left(cov_p\right) + \gamma F \\
s.t. \\
F = \sum_{i=1}^{N} \left\| f_1\left(\left(A \circ A_{binary}\right) \cdot diag(P) \cdot f_2\left(\left(B \circ B_{binary}\right) \cdot diag(U) \cdot u_i + C\right) + D\right) - y_i \right\| \\
A_{binary}, B_{binary}, P, U \in \{0,1\} \\
P_{max} \geqslant \sum_j P_j \geqslant P_{min} \\
A_{binary,ij} \leq P_j \\
B_{binary,ij} \leq U_j \\
A_{LB} \times A_{binary,j} \leq A_{i,j} \leq A_{UB} \times A_{binary,j} \\
B_{LB} \times B_{binary,j} \leq B_{i,j} \leq B_{UB} \times B_{binary,j} \\
C_{LB} \times P_j \leq C_j \leq C_{UB} \times P_j \\
D_{LB} \leq D \leq D_{UB} \\
-4 = A_{LB}, B_{LB}, C_{LB}, D_{LB} \\
4 = A_{UB}, B_{UB}, C_{UB}, D_{UB}
\end{gathered}
\tag{7}
$$

where γ is a tuning parameter to leverage the multi-objective nature of the problem; P_{min} is the minimum number of hidden neurons; P_{max} is the maximum number of hidden neurons. Lower and upper bounds (*LB* and *UB*) of continuous variables are shown using subscripts. These lower and upper bounds are set as -4 and 4, respectively.

$B_{binary,ij}$ is the existence of connection from the j^{th} input as the input information is transferred to i^{th} hidden layer neuron. Once $B_{binary,ij}$ is zero, the connection between j^{th} input and i^{th} neuron is eliminated since no information is transferred due to $B_{binary,ij}$. Thus, once a particular column of B_{binary} is zero; no information from the corresponding input can be transferred to the hidden layer. Corresponding U_j is set as zero by the algorithm. Similarly, once the value of P is zero, the information is not transferred through the corresponding neuron, which therefore means that the neuron is eliminated. In parallel, A_{binary} is the matrix of connection existence between hidden layer and outputs. All these rules are enforced via introducing logic constraints to the formulation in (7).

Equation (7) considers the parameter covariance in the objective function in addition to the training error, where both are highly influenced by the number of parameters and the connections in ANN. From Equation (7), optimal synthesis and training of the corresponding ANN can be employed automatically and simultaneously to obtain OA-ANN. This way, selection of the features and proper conditions are achieved subject to parameter uncertainty manifolds.

There are three types of methods to solve the corresponding mixed-integer type optimization problems, namely evolutionary and derivative-based methods. Two of them can also be combined in a hybrid sense to come up with meta-heuristic method, whose application area has been widening lately [42-44]. Typically, rigorous and derivative-based methods may require

substantial computational power to solve the MINLP problems to global optimality and is out of the scope of this paper. Yet, it must be here noted that suggested formulations bring about the possibility to obtain global ANN structures when solved with non-convex derivative-based methods. Therefore, in this study, an adaptive, evolutionary and heuristic solution algorithm together with a local optimization method is suggested for solving the non-convex MINLP proposed in this paper. Please note that similar adaptive methods described in the previous studies can also be utilized to solve the resulting MINLP problems. This method can also be implemented using open-source codes, which is another vital advantage over using many of the commercial solvers.

As mentioned earlier, an adaptive algorithm is selected and the main aim is to divide the original problem into two parts. Accordingly, the solution is obtained through a hierarchical decomposition of binary and continuous decision variables, as outer and inner loops similar to [45–47]. Accordingly, the outer loop optimization is an integer programming problem (IP) determining existence of the neurons (P), connections (A_{Binary}, B_{Binary}) and selection of input variables (U). After the outer loop is utilized, the inner loop will decide on the optimal weight values for a fixed neural network topology at the current iteration. This decomposition allows faster and effective solution of the original method albeit global optimal cannot be guaranteed.

The outer integer program is given by:

$$\begin{gathered} Min_{A_{binary},B_{binary},P,U} \sum diag\left(cov_p\right) \\ s.t. \\ F \leq \gamma' \\ A_{binary}, B_{binary}, P, U \in \{0,1\} \\ A_{binary,ij} \leq P_j \\ B_{binary,ij} \leq U_j \\ P_{max} \geqslant \sum_i P_j \geqslant P_{min} \end{gathered} \tag{8}$$

where γ' is the maximum desired training error. Note that, the multi objective optimization formulation in Equation (7), is further modified and training error term is considered as a constraint to avoid the difficulty in the determination of γ. In practice, larger value of γ in the solution of Equation (7) may result in over simplification of the model, which in turn causes poor training performance.

The inner loop optimization problem is given by:

$$\min_{A,B,C,D} F$$
$$s.t.$$
$$F = \sum_{i=1}^{N} \left\| f_1\left(\left(A \circ A_{binary}\right) \cdot diag(P) \cdot f_2\left(\left(B \circ B_{binary}\right) \cdot diag(U) \cdot u_i + C\right) + D\right) - y_i \right\|$$
$$A_{LB} \times A_{binary,j} \leq A_{i,j} \leq A_{UB} \times A_{binary,j} \quad (9)$$
$$B_{LB} \times B_{binary,j} \leq B_{i,j} \leq B_{UB} \times B_{binary,j}$$
$$C_{LB} \times P_j \leq C_j \leq C_{UB} \times P_j$$
$$D_{LB} \leq D \leq D_{UB}$$
$$-4 = A_{LB}, B_{LB}, C_{LB}, D_{LB}$$
$$4 = A_{UB}, B_{UB}, C_{UB}, D_{UB}$$

The inner loop is a nonlinear programming problem (NLP), used typically for training a particular architecture iterate given by Problem (8). The IP given in (8) is solved via the MIDACO solver [48,49], whereas (9) is solved via the open-source IPOPT code [50] on an Intel i5 processor with 8GB of RAM running MATLAB 2020a.

Overall heuristic solution algorithm is shown in Fig. 1.

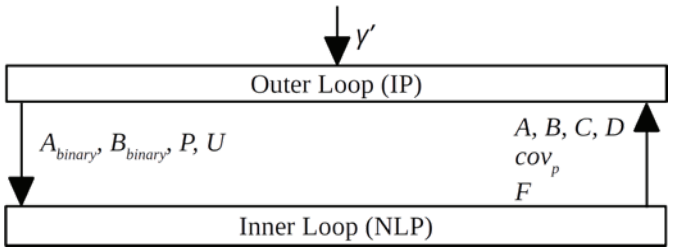

Figure 1. MINLP heuristic solution algorithm

The inner and outer loops are iterated sequentially until the pre-specified convergence (stopping) criterion is reached for the original problem. In this study, this criterion is on the change of the original problem objective, with a value of 0.01. Finally, it is still noteworthy to mention that better solution algorithms can be used while this study mainly focuses on the MINLP approach to combine the feature and structure detection together with the parameter uncertainty prediction for regression problems.

Results

This study focuses on two publicly available benchmarks from [51,52]. The performance of the proposed approach is compared to FC-ANN (Fully

Connected) and several other publications which focus on the same dataset. In addition, GMDH (Group Method of Data Handling) results are also provided using [53]. We decided to compare the proposed method with the GMDH so as to provide a benchmark using an active subject for pruning.

The performances are evaluated using Mean Absolute Error (*MAE*), Mean Square Error (*MSE*), Root Mean Square Error (*RMSE*), Coefficient of Variation (*CV*), and Mean Uncertainty (*MU*), which is a statistical criterion defined in this study, and are calculated from:

$$MU = \frac{\sum_{i=1}^{N} \sqrt{(cov_{y,ii})}}{N}$$
$$MAE = \frac{\sum_{i=1}^{N} \left| y^{prediction,i} - y^{measurement,i} \right|}{N}$$
$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y^{prediction,i} - y^{measurement,i})^2}{N}} \quad (10)$$
$$CV = \frac{\sqrt{\sum_{i=1}^{N} (y^{prediction,i} - y^{measurement,i})^2 / N}}{\sum_{i=1}^{N} y^{prediction,i} / N}$$

where $y^{prediction,i}$ is the prediction of the i^{th} sample output; $y^{measurement,i}$ is the measurement of the i^{th} sample output; N is the number of samples; $cov_{y,ii}$ is the ith diagonal element of cov_y.

Case Study 1

The data set collected by the U.S. Census Service on Boston housing prices and the affecting factors are under consideration [54]. The dataset contains 506 different samples of 13 inputs and single output as shown in Table 1. Randomly selected 50% of the data is used for training and normalized for numerical purposes.

Boston housing dataset is specifically chosen as a case study since it contains relatively low number of samples, and overfitting is highly likely when high number of parameters is introduced. In addition, it has many inputs based on residential and cultural measurements which contain some correlation inherently; and thus input selection and elimination becomes an important issue.

Table 1. Variables of the Case Study 1

			mean	Standard deviation	minimum	maximum
Inputs	1	per capita crime rate by town	3.6	8.6	0.0	89.0
	2	proportion of residential zones for lots over 25K sq.ft.	11.4	23.3	0.0	100.0
	3	proportion of non-retail business acres per town	11.1	6.9	0.5	27.7
	4	Charles River dummy variable	0.1	0.3	0.0	1.0
	5	nitric oxides concentration	0.6	0.1	0.4	0.9
	6	average number of rooms per dwelling	6.3	0.7	3.6	8.8
	7	fraction of owner-occupied units prior to 1940	68.6	28.1	2.9	100.0
	8	weighted distances to five Boston employment centers	3.8	2.1	1.1	12.1
	9	index of accessibility to radial highways	9.5	8.7	1.0	24.0
	10	full-value property-tax rate per $10,000	408.2	168.5	187.0	711.0
	11	pupil-teacher ratio by town	18.5	2.2	12.6	22.0
	12	1000(Bk - 0.63)2 where Bk is the proportion of a particular resident group in town	356.7	91.3	0.3	396.9
	13	% lower status of the population	12.7	7.1	1.7	38.0
Output		Median value of owner-occupied homes in $1000s	22.5	9.2	5.0	50.0

Fig. 2 includes training and test performances of a FC-ANN containing 10 neurons in the hidden layer. The FC-ANN contains 151 continuous parameters to be estimated. Accurate estimation of such a high number of parameters is theoretically challenging and likely to result in overfitting, considering 253 training samples with 13 inputs.

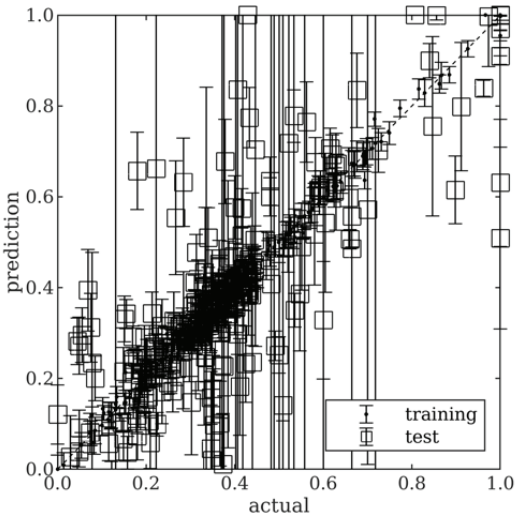

Figure 2. The performance of FC-ANN with 10 neurons

FC-ANN delivers a relatively better training performance due to high number of connections, neurons and inputs. On the other hand, a significant performance drop is observed for the test data due to overfitting, in this particular relatively smaller case study. The error bars of the predictions are obtained from (5), using the uncertainties in the parameters after the training. Please note that these measures could represent prediction robustness and reliability. (6) delivers the mean value of the predictions based on the mean parameter values at a particular architecture. Due to probable overfitting in FC-ANN, the prediction uncertainty and error are significantly large, which in practice means predictions might not reliable.

Solutions of (8) and (9) deliver the OA-ANN, whose performance is shown in Fig. 3.

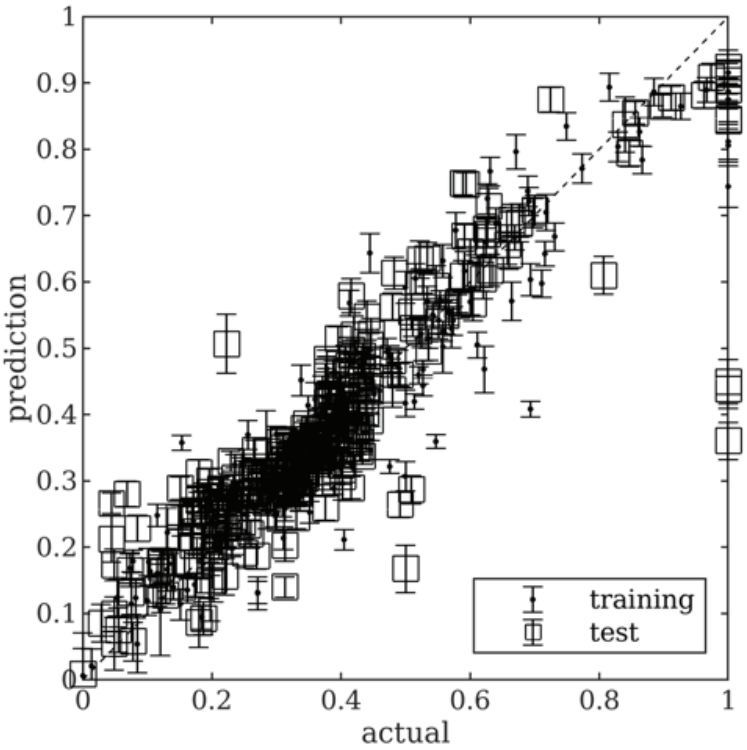

Figure 3. The performance of OA-ANN with 2 neurons

Table 2 provides detailed statistical comparisons based on the common measures. The performance increase is obtained through optimal architecture design and training. Note that all OA-ANN test errors and prediction uncertainty range are lower than FC-ANN as shown in Table 2, in this particular case. Corresponding OA-ANN architecture, which does not explicitly demonstrate the bias connections, is shown in Fig. 4.

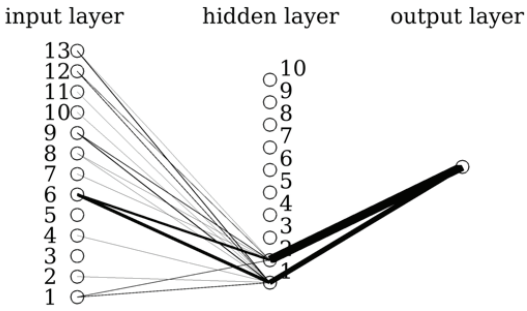

Figure 4. OA-ANN structure and connections for case study 1

As shown in Fig. 4, in this particular case, only 2 neurons (although maximum 10 is allowed) are introduced with the elimination of proportion of non-retail business acres per town and nitric oxides concentration from the input set. Note that there is no connection between the corresponding input and any hidden neuron. In addition, OA-ANN contains a significantly fewer number of connections among variables; for instance, Charles River dummy variable provides information in the calculation of the output variable, through the 1st hidden neuron only. Furthermore, the connection line widths are scaled by the absolute values of the corresponding weight.

Table 2. Results of Case Study 1

		OA-ANN	FC-ANN	[55]	[56]	GMDH
training	*MAE*	0.050	0.013	-	-	-
	MSE	0.004	0.0003	-	-	0.006
	RMSE	0.067	0.019	-	-	0.078
	*RMSE**	3.086	0.849	-	3.369	-
	MU	0.018	0.278	-	-	-
test	*MAE*	0.058	0.097	-	-	-
	MSE	0.009	0.024	-	-	0.012
	RMSE	0.096	0.155	0.187	-	0.119
	*RMSE**	4.331	6.990	-	7.602	-
	MU	0.018	0.195	-	-	-
$\sum diag\,(cov_p)$		0.533	27998	-	-	-
Neuron		2	10	-	20	8
Inputs		11	13	13	6	6
Connections		22	151	-	-	65

*Calculated without data normalization.

The OA-ANN, using fewer network elements, provides a comparable performance with the benchmarks in the literature and GMDH. In [55], an extreme learning machine confidence weighted method is proposed using 79% of the whole data

in training. [56] used 60% of the whole data, and reported radial basis neural network results using different number of neurons. [57] also refers to various other models and provides a performance comparison on Boston housing dataset with test RMSE* values between 3.206 and 7.610. In our particular case, OA-ANN has better performance compared to most of the other approaches.

Case Study 2

Our second case study is related to predicting the electricity consumption of a building. The dataset includes relatively higher number of samples, being with 4208 points, and is directly taken from PROBEN 1 benchmark problem set [52]. The dimension of the input vector is 14 in total as some of the inputs are lumped into each other in the original dataset [58]. The electricity consumption (WBE) is predicted based on year, month, date, day of the week, time of day, outside temperature, outside air humidity, solar radiation and wind speed. The statistical description of the dataset is summarized in Table 3.

Table 3. Statistical description of Case Study 2

	Inputs														Output
Mean	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.5	0.0	0.0	0.6	0.3	0.1	0.2	0.3
Standard Deviation	0.4	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.6	1.0	0.2	0.2	0.2	0.1	0.1
Minimum	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-1.0	-1.0	0.0	0.0	0.0	0.0	0.1
Maximum	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.9	0.7

Performance of the proposed OA-ANN on the prediction of building electricity consumption is compared with fully-connected ANN (FC-ANN) and GMDH, and two well-known benchmarks using the same dataset taken from the literature [59]. In [59], the authors used single hidden layer feedforward ANN structure employing hyperbolic tangent activation functions. They introduced identification, additive and subtractive phases into their training algorithm as opposed to classical methods and sequentially analyzed the effects of the number of inputs and the number of neurons. 70 % of the whole data is used for all phases described in the paper. Results showed that reducing the geometry of the ANNs could yield much better test results.

In [60], a hybrid genetic algorithm-adaptive network based fuzzy inference system is proposed to train feedforward artificial neural networks. 70% of the whole data is used for training. This paper also proposes an optimization-based training method and reduces the size of the ANNs using sequential

analysis. However, this method is not an automatic and simultaneous design and training method and does not take the covariance of parameters into account.

The proposed method is implemented with 15 maximum number of hidden neurons (P_{max}), using 70% of the whole data for training. Same training and test dataset, input and neuron numbers are implemented to FC-ANN and GMDH for fair and clear comparison. All results are reported in Table 4.

Results for OA-ANN show that the method proposes to use 6 inputs, being year, month, type of day, temperature, solar radiation and wind speed.

Table 4. Results of Case Study 2

		OA-ANN	FC-ANN	[60]	[61]	GMDH
training	*MSE*	0.0026	0.0009	0.008	-	-
	MAE	0.042	0.038	-	-	-
	CV	12	10	12	9.6	11.3
	RMSE	0.051	0.03	0.09	-	0.07
	MU	0.0043	0.015	-	-	-
test	*MSE*	0.002	0.0043	0.02	-	-
	MAE	0.041	0.06	-	-	-
	CV	9.3	11	13	10	11.1
	RMSE	0.046	0.065	0.14	-	0.069
	MU	0.0044	0.016	-	-	-
$\sum diag\,(cov_p)$		2.8	1.2×10^{6}	-	-	-
Neuron		5	15	4	4	6
Inputs		6	14	10	5	7
Connections		23	241	49	29	55

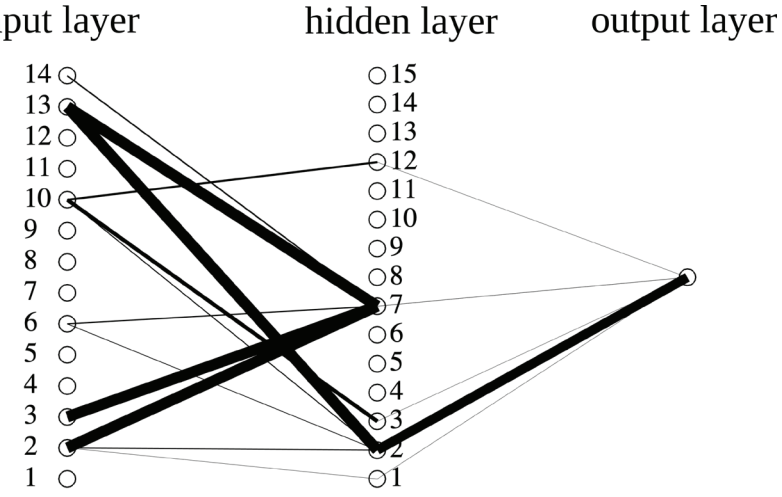

Figure 5. OA-ANN structure and connections for case study 2

As shown in Fig. 5, in this particular case, 5 neurons (out of 15) are selected. Similar to first case study, the connection line widths are scaled by the absolute values of the corresponding weight.

Table 4 provides detailed statistical comparisons based on the common measures. Coefficient of Variation (*CV*), which is used in other benchmarks, is also included into this case study [60]. Table 4 shows that OA-ANN includes 23 connections among all variables in total, which is significantly fewer than the FC-ANN. Accordingly, OA-ANN exhibits poor training performance than the FC-ANN in all statistical metrics except mean uncertainty. On the other hand, OA-ANN provides better test performance than the traditional FC-ANN in spite of using fewer numbers of inputs, neurons and connections. Such improved prediction quality increase is obtained with almost 90% decrease in the number of connections compared to FC-ANN.

Similarly, the benefit of size reduction and pruning for ANNs in the context of optimization can be observed from [60], whose performance is relatively better compared to other benchmark studies using the same dataset [60]. Even though OA-ANN has poorer training performance compared to [60], OA-ANN exhibits the best test performance among all benchmarks reported in this paper, both in terms of reduced standard deviation and test error. The main reason for this observation is the fact that OA-ANN considers the covariance of the parameters as an optimization metric to be minimized, in addition to the training objective function. As a result, uncertainty regions of the OA-ANN predictions are tightened, as shown in Table III. This tightening ultimately brings about much fewer values for *MU, CV* and $\sum diag\left(cov_p\right)$, which in turn enhances both accuracy and precision of model predictions.

Conclusions

This study focuses on the simultaneous optimal architecture ANN design and training algorithm under parameter uncertainty and uncertainty propagation considerations for regression problems; in contrast to traditional approaches where the structure is fixed by predefined hyper parameters based on trial and error procedure. The existence of the connections, the selection of input variables and the determination of the number of hidden neurons together with connection weights are under consideration. The main aim of this formulation is to obtain the optimal ANN structure, and to train this structure with the most dependable input variables considering the parameter identifiability issues to deliver a prediction with lower confidence interval, or to be more precise, uncertainty.

The proposed approach integrates the design and training simultaneously through an MINLP problem which is decomposed for the utilization in

successively solved smaller optimization problems, mainly IP and NLP. The MINLP problem involves the training error and the parameter covariance matrix as an uncertainty measure. It also ensures the selection of identifiable set of parameters, resulting in a more robust prediction performance. Furthermore, the MINLP problem includes extra logic constrains for a more efficient solver performance.

The proposed MINLP formulation is comprehensive, sophisticated and modifiable to other ANN types (i.e. recurrent ANNs, Convolutional Neural Networks). However, similar to many machine learning algorithms, ANNs suffer from nonconvex optimization problem due to nonlinear activation functions and the performance is highly sensitive to initial guess and optimization algorithm, which might deliver local optimum with different ANN weights despite processing same training data. The proposed formulation further increases the complexity of the training problem by introducing binary variables to represent the existence of a particular network element. Such a desirable theoretical superiority calls for mixed-integer optimization algorithms whose global optimum finding capability is still limited with a complex nonconvex problem and require significant computational load due to rigorous formulation. On the other hand, convex MINLP solvers, similar to heuristic optimization algorithms, deliver a local optimum with a significant and computation time increases drastically since all variables are modified simultaneously in the iterations. For such considerations, in this study, a pseudo-decomposition is applied to obtain a satisfactory ANN architecture and performance through computationally favorable heuristic method. The proposed heuristic solution method also suffers from local optimality issues since no explicit modification is implemented to handle nonconvexity related problems. The integer programming stage in the nested algorithm enables the evaluation of any blackbox formulation in the inner loop; but makes the overall solution exposed to failures once the tuning of the corresponding stage optimization problem is poor or not compatible with the inner loop. In addition, the interactions of the layers might bring additional infeasibility problems since there the problems process different constraints. Some linking constraints are introduced to tighten the search space and bring computational efficiency. However, the proposed pseudo-decomposition benefits from rigorous formulations in the inner loop where nonlinear programming is performed with sophisticated mathematical developments including algorithmic differentiation. The development of pseudo-decomposition through using more sophisticated optimization algorithms with a better tuning combined with feasibility cuts and pumps would further increase the computational efficiency, which is under consideration for our future works.

The proposed approach is implemented on two publicly available datasets which are studied extensively in the literature. It is shown that, the current

approach provides a better test performance despite increased training error. Finally, the current research activities focus on extending the suggested framework for deep and recurrent neural networks and for synthesizing more efficient neural network based controller structures.

Acknowledgements: The authors thank to Prof. Dr. Taskin Kavzoglu for his valuable comments and discussions.

Funding Information: EA acknowledges research funding from TUBITAK 2232 Program. This publication has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C245). However, the entire responsibility of the publication belongs to the owner of the publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Llaneras, F.; Sala, A.; Picó, J. Dynamic Estimations of Metabolic Fluxes with Constraint-Based Models and Possibility Theory. *J. Process Control* 2012, *22*, 1946–1955, doi:10.1016/j.jprocont.2012.08.016.
2. McCulloch, W.S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* 1943, *5*, 115–133, doi:10.1007/BF02478259.
3. Vellido, A.; Lisboa, P.J.G.; Vaughan, J. Neural Networks in Business: A Survey of Applications (1992-1998). *Expert Syst. Appl.* 1999, *17*, 51–70, doi:10.1016/s0957-4174(99)00016-0.
4. Park, B.; Kwon Bae, J. Using Machine Learning Algorithms for Housing Price Prediction: The Case of Fairfax County, Virginia Housing Data. *Expert Syst. Appl.* 2015, *42*, 2928–2934, doi:10.1016/j.eswa.2014.11.040.
5. Venkatasubramanian, V. The Promise of Artificial Intelligence in Chemical Engineering: Is It Here, Finally? *AIChE J.* 2019, *65*, 466–478, doi:10.1002/aic.16489.
6. Lee, J.H.; Shin, J.; Realff, M.J. Machine Learning: Overview of the Recent Progresses and Implications for the Process Systems Engineering Field. *Comput. Chem. Eng.* 2018, *114*, 111–121, doi:10.1016/j.compchemeng.2017.10.008.
7. Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.E.; Arshad, H. State-of-the-Art in Artificial Neural Network Applications: A Survey. *Heliyon* 2018, *4*.
8. Qin, S.J.; Chiang, L.H. Advances and Opportunities in Machine Learning for Process Data Analytics. *Comput. Chem. Eng.* 2019, 465–473, doi:10.1016/j.compchemeng.2019.04.003.
9. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Hasan, M.; Van Essen, B.C.; Awwal, A.A.S.; Asari, V.K. A State-of-the-Art Survey on Deep Learning Theory and Architectures. *Electronics* 2019, *8*, 292, doi:10.3390/electronics8030292.
10. Murugan, C.; Natarajan, P. Estimation of Fungal Biomass Using Multiphase Artificial Neural Network Based Dynamic Soft Sensor. *J. Microbiol. Methods* 2019, *159*, 5–11, doi:10.1016/j.mimet.2019.02.002.
11. Kaur, H.; Kumari, V. Predictive Modelling and Analytics for Diabetes Using a Machine Learning Approach. *Appl. Comput. Informatics* 2019, doi:10.1016/j.aci.2018.12.004.
12. McCoy, J.T.; Auret, L. Machine Learning Applications in Minerals Processing: A Review. *Miner. Eng.* 2019, *132*, 95–109, doi:10.1016/j.mineng.2018.12.004.

13. Ghoddusi, H.; Creamer, G.G.; Rafizadeh, N. Machine Learning in Energy Economics and Finance: A Review. *Energy Econ.* 2019, *81*, 709–727, doi:10.1016/j.eneco.2019.05.006.
14. Acheampong, A.O.; Boateng, E.B. Modelling Carbon Emission Intensity: Application of Artificial Neural Network. *J. Clean. Prod.* 2019, *225*, 833–856, doi:10.1016/j.jclepro.2019.03.352.
15. Cabaneros, S.M.; Calautit, J.K.; Hughes, B.R. A Review of Artificial Neural Network Models for Ambient Air Pollution Prediction. *Environ. Model. Softw.* 2019, *119*, 285–304, doi:10.1016/j.envsoft.2019.06.014.
16. Li, L.; Gao, Z.; Yuan, Z.M. On the Sensitivity and Uncertainty of Wave Energy Conversion with an Artificial Neural-Network-Based Controller. *Ocean Eng.* 2019, *183*, 282–293, doi:10.1016/j.oceaneng.2019.05.003.
17. Macmurray, J.C.; Himmelblau, D.M. Modeling and Control of a Packed Distillation Column Using Artificial Neural Networks. *Comput. Chem. Eng.* 1995, *19*, 1077–1088, doi:10.1016/0098-1354(94)00098-9.
18. Biswas, M.A.R.; Robinson, M.D.; Fumo, N. Prediction of Residential Building Energy Consumption: A Neural Network Approach. *Energy* 2016, *117*, 84–92, doi:10.1016/j.energy.2016.10.066.
19. Ma, J.; Xia, D.; Wang, Y.; Niu, X; Jiang, S.; Liu, Z; Guo, H. A comprehensive comparison among metaheuristics (MHs) for geohazard modeling using machine learning: Insights from a case study of landslide displacement prediction. *Engineering Applications of Artificial Intelligence* 2022, 114, 105150, doi: 10.1016/j.engappai.2022.105150.
20. Ma, J.; Xia, D.; Guo, H.; Wang, Y.; Niu, X; Liu, Z; Jiang, S. Metaheuristic-based support vector regression for landslide displacement prediction: a comparative study. *Landsides* 2022, 1-23, doi: 0.1007/s10346-022-01923-6
21. Ardiansyah, A.; Naufalin, R.; Arsil, P.; Latifasari, N.; Wicaksono, R.; Aliim, M.S.; Kartiko, C.; Waluyo, S. Machine Learning Model for Quality Parameters Prediction and Control System Design in the Kecombrang Flower (Etlingera Elatior) Extraction Process. *Process.* 2022, *10*.
22. Levasseur, L.P.; Hezaveh, Y.D.; Wechsler, R.H. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *arXiv* 2017, *850*, L7, doi:10.3847/2041-8213/aa9704.
23. Schittkowski, K. Experimental Design Tools for Ordinary and Algebraic Differential Equations. In Proceedings of the Industrial and Engineering Chemistry Research; December 2007; Vol. 46, pp. 9137–9147.
24. Fernández, J.; Chiachío, M.; Chiachío, J.; Muñoz, R.; Herrera, F. Uncertainty Quantification in Neural Networks by Approximate Bayesian Computation: Application to Fatigue in Composite Materials. *Eng. Appl. Artif. Intell.* 2022, *107*, 104511, doi:10.1016/j.engappai.2021.104511.
25. Ivakhnenko, a. G. The Group Method of Data Handling-a Rival of the Method of Stochastic Approximation. *Sov. Autom. Control* 1968, *13*, 43–71.
26. Yang, L.; Yang, H.; Yang, H.; Liu, H. GMDH-Based Semi-Supervised Feature Selection for Electricity Load Classification Forecasting. *Sustain.* 2018, *10*, doi:10.3390/su10010217.
27. Ivakhnenko, A.G. Sorting Methods in Self-Organization of Models and Clusterizations (Review of New Basic Ideas) Iterative (Multirow) Polynomial GMDH Algorithms. *Sov. J. Autom. Inf. Sci.* 1989, *22*, 88–99.
28. Dua, V. A Mixed-Integer Programming Approach for Optimal Configuration of Artificial Neural Networks. *Chem. Eng. Res. Des.* 2010, *88*, 55–60, doi:10.1016/j.cherd.2009.06.007.
29. Sildir, H.; Aydin, E.; Kavzoglu, T. Design of Feedforward Neural Networks in the Classification of Hyperspectral Imagery Using Superstructural Optimization. *Remote Sens.* 2020, *12*, doi:10.3390/rs12060956.
30. Sildir, H.; Aydin, E. A Mixed-Integer Linear Programming Based Training and Feature Selection Method for Artificial Neural Networks Using Piece-Wise Linear Approximations. *Chem. Eng. Sci.* 2022, *249*, 117273, doi:10.1016/j.ces.2021.117273.

31. Sildir, H.; Sarrafi, S.; Aydin, E. Data-Driven Modeling of an Industrial Ethylene Oxide Plant: Superstructure-Based Optimal Design for Artificial Neural Networks. In *Computer Aided Chemical Engineering*; Elsevier, 2021; Vol. 50, pp. 445–450 ISBN 1570-7946.
32. Sildir, H.; Sarrafi, S.; Aydin, E. Optimal Artificial Neural Network Architecture Design for Modeling an Industrial Ethylene Oxide Plant. *Comput. Chem. Eng.* 2022, *163*, 107850, doi:https://doi.org/10.1016/j.compchemeng.2022.107850.
33. Feng, J.; Chen, J.; Sun, Q.; Shang, R.; Cao, X.; Zhang, X.; Jiao, L. Convolutional Neural Network Based on Bandwise-Independent Convolution and Hard Thresholding for Hyperspectral Band Selection. *IEEE Trans. Cybern.* 2020, 1–15, doi:10.1109/TCYB.2020.3000725.
34. Piotrowski, A.P.; Napiorkowski, J.J. A Comparison of Methods to Avoid Overfitting in Neural Networks Training in the Case of Catchment Runoff Modelling. *J. Hydrol.* 2013, *476*, 97–111, doi:10.1016/j.jhydrol.2012.10.019.
35. Sun, Y.; Xue, B.; Zhang, M.; Yen, G.G.; Lv, J. Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification. *IEEE Trans. Cybern.* 2020, doi:10.1109/tcyb.2020.2983860.
36. Benardos, P.G.; Vosniakos, G.C. Optimizing Feedforward Artificial Neural Network Architecture. *Eng. Appl. Artif. Intell.* 2007, *20*, 365–382, doi:10.1016/j.engappai.2006.06.005.
37. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Networks* 2015, *61*, 85–117.
38. Mclean, K.A.P.; Mcauley, K.B. Mathematical Modelling of Chemical Processes-Obtaining the Best Model Predictions and Parameter Estimates Using Identifiability and Estimability Procedures. *Can. J. Chem. Eng.* 2012, *90*, 351–366, doi:10.1002/cjce.20660.
39. Lin, Z.; Zou, Q.; Ward, E.S.; Ober, R.J. Cramer-Rao Lower Bound for Parameter Estimation in Nonlinear Systems. *IEEE Signal Process. Lett.* 2005, *12*, 855–858, doi:10.1109/LSP.2005.859498.
40. Tellinghuisen, J. Statistical Error Propagation. *J. Phys. Chem. A* 2001, *105*, 3917–3921, doi:10.1021/jp003484u.
41. Tirkolaee, B.E.; Goli, A.; Weber, G. W. Fuzzy Mathematical Programming and Self-Adaptive Artificial Fish Swarm Algorithm for Just-in-Time Energy-Aware Flow Shop Scheduling Problem With Outsourcing Option. *IEEE Transactions on Fuzzy Systems* 2020, 28, 2772-2783, doi: 10.1109/TFUZZ.2020.2998174
42. Goli, A.; Tirkolaee, E.B.; Sangaiah, A.K. Hybrid neural network and improved cuckoo optimization algorithm for forecasting thermal comfort index at urban open spaces. *Advances in Edge Computing: Massive Parallel Processing and Applications* 2020, 264, IOS Press.
43. Goli, A.; Khademi-Zare, H.; Tavakkoli-Moghaddam, R.; Sadeghieh, A.; Sasanian, M.; Malekalipour Kordestanizadeh, R. An integrated approach based on artificial intelligence and novel meta-heuristic algorithms to predict demand for dairy products: a case study. *Network: computation in neural systems* 2021, 32, 1-35, 10.1080/0954898X.2020.1849841
44. Abdel-Aal, R.E. GMDH-Based Feature Ranking and Selection for Improved Classification of Medical Data. *J. Biomed. Inform.* 2005, *38*, 456–468, doi:10.1016/j.jbi.2005.03.003.
45. Pintarič, Z.N.; Kravanja, Z. The Two-Level Strategy for MINLP Synthesis of Process Flowsheets under Uncertainty. In Proceedings of the Computers and Chemical Engineering; July 2000; Vol. 24, pp. 195–201.
46. Chen, X.; Li, Z.; Yang, J.; Shao, Z.; Zhu, L. Nested Tabu Search (TS) and Sequential Quadratic Programming (SQP) Method, Combined with Adaptive Model Reformulation for Heat Exchanger Network Synthesis (HENS). *Ind. Eng. Chem. Res.* 2008, *47*, 2320–2330, doi:10.1021/ie071245o.
47. Chen, X.; Li, Z.; Wan, W.; Zhu, L.; Shao, Z. A Master-Slave Solving Method with Adaptive Model Reformulation Technique for Water Network Synthesis Using MINLP. *Sep. Purif. Technol.* 2012, *98*, 516–530, doi:10.1016/j.seppur.2012.06.039.
48. Schlüter, M.; Gerdts, M.; Rückmann, J.J. A Numerical Study of MIDACO on 100 MINLP Benchmarks. *Optimization* 2012, *61*, 873–900, doi:10.1080/02331934.2012.668545.

49. Schlueter, M. MIDACO Software Performance on Interplanetary Trajectory Benchmarks. *Adv. Sp. Res.* 2014, *54*, 744–754, doi:10.1016/j.asr.2014.05.002.
50. T. Biegler, L. Large-Scale Nonlinear Programming: An Integrating Framework for Enterprise-Wide Dynamic Optimization. *Comput. Aided Chem. Eng.* 2007, *24*, 575–582, doi:10.1016/S1570-7946(07)80119-2.
51. Olson, R.S.; La Cava, W.; Orzechowski, P.; Urbanowicz, R.J.; Moore, J.H. PMLB: A Large Benchmark Suite for Machine Learning Evaluation and Comparison. *BioData Min.* 2017, *10*, 36, doi:10.1186/s13040-017-0154-4.
52. Prechelt, L. Proben1: A Set of Neural Network Benchmark Problems and Benchmarking Rules; Karlsruhe, 1994; Vol. 21;.
53. GMDH Group Method of Data Handling Available online: http://www.gmdh.net/ (accessed on 23 March 2020).
54. Harrison, D.; Rubinfeld, D.L. Hedonic Housing Prices and the Demand for Clean Air. *J. Environ. Econ. Manage.* 1978, *5*, 81–102, doi:10.1016/0095-0696(78)90006-2.
55. Shang, Z.; He, J. Confidence-Weighted Extreme Learning Machine for Regression Problems. *Neurocomputing* 2015, *148*, 544–550, doi:10.1016/j.neucom.2014.07.009.
56. Tsekouras, G.E.; Tsimikas, J. On Training RBF Neural Networks Using Input-Output Fuzzy Clustering and Particle Swarm Optimization. *Fuzzy Sets Syst.* 2013, *221*, 65–89, doi:10.1016/j.fss.2012.10.004.
57. Kim, E.H.; Oh, S.K.; Pedrycz, W. Design of Double Fuzzy Clustering-Driven Context Neural Networks. *Neural Networks* 2018, *104*, 1–14, doi:10.1016/j.neunet.2018.03.018.
58. Mira, J.; Álvarez, J.R. Computational Methods in Neural Modeling: 7th International Work-Conference on Artificial and Natural Neural Networks, IWANN 2003, Maó, Menorca, Spain, June 3-6. Proceedings; Springer Science & Business Media, 2003; Vol. 1; ISBN 9783540402107.
59. Karatasou, S.; Santamouris, M.; Geros, V. Modeling and Predicting Building's Energy Use with Artificial Neural Networks: Methods and Results. *Energy Build.* 2006, *38*, 949–958, doi:10.1016/j.enbuild.2005.11.005.
60. Li, K.; Su, H.; Chu, J. Forecasting Building Energy Consumption Using Neural Networks and Hybrid Neuro-Fuzzy System: A Comparative Study. *Energy Build.* 2011, *43*, 2893–2899, doi:10.1016/j.enbuild.2011.07.010.

4
Superstructure Optimization of Dimethyl Ether Process

Emrullah Erturk[a], Erdal Aydın[b,c], Sahin Sarrafi[d], Ozgun Deliismail[d], Aysel Zahidova[d], Hasan Sildir[a,*]

[a]*Department of Chemical Engineering, Gebze Technical University, Kocaeli 41400, Türkiye*
[b]*Department of Chemical and Biological Engineering, Koç University, Istanbul 34457, Türkiye*
[c]*Koç University TUPRAS Energy Center (KUTEM), Koç University, Istanbul 34450, Türkiye*
[d]*SOCAR Türkiye, R&D and Innovation Co., Aliaga, Izmir, 35800, Türkiye*
[*]*hasansildir@gtu.edu.tr*

Conferance: European Symposium on Computer-Aided Process Engineering-32, Toulouse. https://www.conftool.net/escape32-2022/sessions.php

Journal: Computer Aided Chemical Engineering, Elsevier, Volume 51, **2022**, 661-666

https://www.sciencedirect.com/science/article/abs/pii/B9780323958790501119

DOI: https://doi.org/10.1016/B978-0-323-95879-0.50111-9

Abstract

Integration of process flowsheet simulators and optimization algorithms is a prominent approach to address simultaneous design and optimization of processes, which is represented by a mixed integer nonlinear programming (MINLP) formulation. In this study, DWSIM, a free and rarely used simulator, is used as a black-box function for the evaluation in genetic algorithm in MATLAB. Proposed approach is implemented to a dimethyl ether process, calculating optimum processing conditions in addition to structural decision variables including the feedstock type, reactor, and separation unit selections. Results show that syngas has the major impact on the process economics and is significantly more economical feedstock although high number of additional processing units are required.

Keywords: superstructure optimization; mixed integer nonlinear programming; process synthesis.

Introduction

Chemical processes include complex and integrated pathways combined by several unit operations in which conversion occurs from feedstocks to valuable products. Such a conversion, in general, can be obtained through high number of alternative paths, which hinders the process architecture and operating condition selection under economic considerations and tightly ensured process constraints. Thus, a superstructure optimization problem formulation

addressing simultaneously the aforementioned issues has become an important research area over the past decades.

A mixed-integer nonlinear problem (MINLP) is formulated, in general, to obtain a smaller architecture from a larger superstructure that is predefined and includes all alternative units, flows and many other considerations. The resulting MINLP formulation includes integer variables to account for sequences of events, alternative candidates, and the existence of units, whereas the continuous variables represent states. The MINLP problem is flexible and can include various user-defined considerations from economic and environmental aspects [1].

Significant developments in theory and computation have provided substantial capabilities to solve various complex optimization problems over the past decades [2]. With many alternatives and modifications, MINLP solvers can be classified roughly into two major groups as rigorous, and heuristic based on solution approach. Rigorous solvers utilize all mathematical tools to exploit optimality through the explicit formulation of all superstructure elements and might ensure global optimality for many cases. As an alternative, black-box or evolutionary algorithms benefit from the patterns from successive function evaluations, rather than focusing on derivative-based iterations, to provide optimality. Evolutionary algorithms, although suffering from significant computational load and local optimality issues, might deliver practical and satisfactory solutions to complex MINLPs once the associated plant models are challenging to derive and requires a priori knowledge due to numerous unknown driving forces and their nonlinearly interacting nature [3]. In contrast, despite significant computational developments and theoretical advancements, rigorous solvers have scalability problems for large superstructures. In addition, integration of those to process flowsheet simulators which have widespread industrial use is a challenging task, hindering rigorous solver implementations. A significant fraction of superstructure optimization studies focus on commercial software [4]. On the other hand, some free alternatives, including DWSIM, might provide similar performance to commercial ones [5]. Teerapat and Amata used a DWSIM simulation environment to validate an optimized DME process [6].

This study focuses on the simultaneous design and optimization of a comprehensive DME process under several scenarios including different feedstock and unit operation prices using a free process flowsheet simulator, DWSIM, which is automated through Python. The superstructure involves different feedstocks which are processed through different reactors and separation units to be determined by the optimization problem in addition to some operating conditions. A genetic algorithm is used for the evolutionary solution of MINLP problem, which accounts for the existence of the units

and plant operating conditions. The case studies show different process configurations are required for profitability based on feedstock and product prices.

Superstructure of DME Production Process

Dimethyl ether is one of the prominent candidates for relatively clean and sustainable alternatives to fossil fuels. The increased demand for DME can be met from synthesis gas or methanol from biological waste [7].

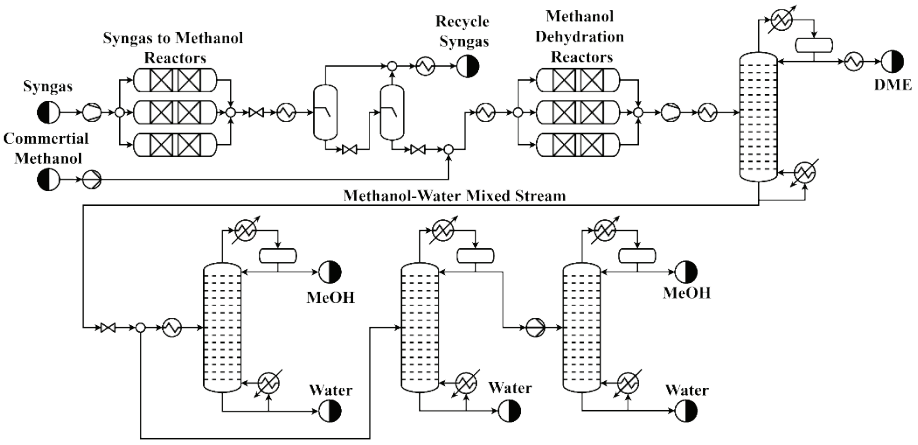

Figure 1. Superstructure of DME production process

The proposed superstructure in this study is presented in Fig. 1. DME production might be obtained with dehydration of methanol and from various feedstocks, primarily from syngas, as an emerging trend to favor carbon management and biomass utilization. The conditions specified in the kinetic study of Nestler et al. were used for modelling methanol synthesis at three different catalyst density yields high to low conversion [8]. Sequential flash units separate syngas from methanol with minimal loss at different pressures. Next, DME is produced with preheated methanol in catalytic dehydration reactor with three conversion preferences. Pressurized and cooled reactor effluent consists of mainly water, methanol and DME, which is obtained using 30 Bar distillation column at high purity. Next, unreacted methanol and side product water recovered with two different distillation column arrangements. Side products are marketable/ recyclable as well.

MINLP Formulation and Solution Method

Superstructure of DME process optimized with MINLP problem formulated as:

$$
\begin{aligned}
&\min_{m,y,T,P} && IC + (OC - S \cdot (1-f)) \cdot t_f \\
&s.t. && y_1 + y_2 = 1 \\
& && y_3 + y_4 + y_5 = 1 \\
& && y_6 + y_7 + y_8 = 1 \\
& && y_9 + y_{10} = 1 \\
& && m_6 = y_6 \cdot (m_1 y_1 + m_{F2} y_2) \\
& && m_7 = y_7 \cdot (m_1 y_1 + m_{F2} y_2) \\
& && m_8 = y_8 \cdot (m_1 y_1 + m_{F2} y_2) \\
& && m_9 = y_9 \cdot m_{T101,B} \\
& && m_{10} = y_{10} \cdot m_{T101,B} \\
& && y_2 \cdot y_3 \cdot 6 \leq m_3 \leq y_2 \cdot y_3 \cdot 10.5 \\
& && y_2 \cdot y_4 \cdot 6 \leq m_4 \leq y_2 \cdot y_4 \cdot 12.5 \\
& && y_2 \cdot y_5 \cdot 6 \leq m_5 \leq y_2 \cdot y_5 \cdot 15.5 \\
& && y_1 \cdot m_{1,min} \leq m_1 \leq y_1 \cdot m_{1,max} \\
& && y_1 \cdot m_{2,min} \leq m_2 \leq y_2 \cdot m_{2,max} \\
& && 301 \leq T_1 \leq 308 \\
& && 360 \leq T_2 \leq 380 \\
& && 46 \leq P_1 \leq 50 \\
& && y_j \in \{0,1\}\ j = 1 \ldots 10
\end{aligned}
\tag{1}
$$

where t_f is the project lifetime which is 20 years for this case; f is the tax rate; y_1 and y_2 are binary variables for commercial methanol and syngas feedstock selection, respectively; y_3, y_4, and y_5 represent the selection of syngas to methanol reactor with corresponding mass flow rates m_3, m_4 and m_5; y_6, y_7, and y_8 methanol dehydration reactor selection variables; m_{F2} is the flash unit effluent; $m_{T101,B}$ is the bottom product of DME distillation column; y_9 and y_{10} are one column and sequential two-column distillation route selection binaries, respectively; S is the income from the products. Equality constraints which include binary variables solely ensure selection of a single path among many. Commercial methanol and syngas stream mass flow rates are represented with m_1 and m_2. High to low DME reactor inlet stream mass flow rates are m_6, m_7 and m_8, respectively. m_9 is 30-stage column and m_{10} is sequential two 15-stage columns mass flow rates. $m_{F102,B}$ is second flash unit bottom product stream that contains mainly produced methanol, mass flow rate. All produced or purchased methanol fed to DME reactor and this feed stream, $m_{R2,in}$, expressed with fifth equality constraint. $m_{T101,B}$ is bottom product mass flow rate of DME recovery column which fed to selected separation route. T_1, P_1 are temperature and pressure of flash unit, and T_2 DME recovery column feed stream temperature, respectively. Inequality constraints limit syngas feed mass flow for prevent exceeding equipment sizes which assumed fixed for practical calculation of

capital cost investments. The capital cost investment, IC, the operating cost, OC, is calculated from:

$$IC = F_{LANG} \cdot \left(\sum_i C_{BM,i} \cdot y_i + \sum_j C_{BM,fixed,j}\right)$$
$$OC = m_1 \cdot y_1 \cdot C_{MeOH} + m_2 \cdot y_2 \cdot C_{Syngas} + \sum_j Q_j C_{utility,j} \quad (2)$$

where F_{LANG} is the correction for the liquid phase operations; $C_{BM,i}$ is the bare module cost of i^{th} equipment which is related to the y_i; $C_{BM,j}$ is the bare module cost of j^{th} equipment which exist in the structure independent of the optimization problem; C_{MeOH} and C_{Syngas} are methanol and syngas prices, respectively; Q_j and $C_{utility,j}$ is the quantity and the unit cost of j^{th} utility. Equipment purchases and operational cost calculations performed using [9] and updated to current prices using CEPCI. The distillation column fixed cost provided from ChemSep. Process equipment sizing considers the dimensions that could meet the highest production capacity. For this reason, even if there is a high investment cost in low-capacity production preferences, meeting the upper limit production capacity is ensured.

The MINLP problem in Eq. 1 is solved using Genetic Algorithm (GA) which treats the process flow sheet simulation as a black-box function which is evaluated in DWSIM. Python has a key role in the implementation of the MINLP problem as it provides the communication between DWSIM and MATLAB's GA. A simplified information flow diagram is shown in Fig. 2.

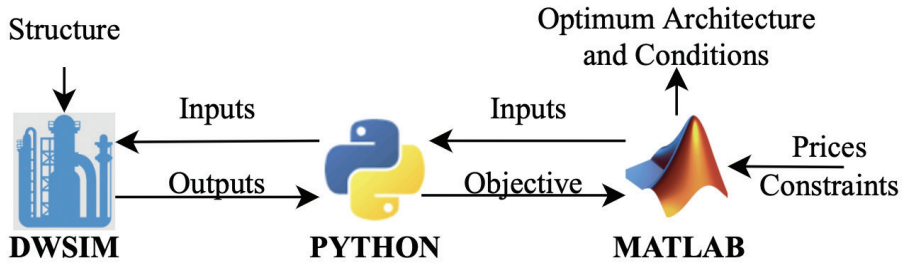

Figure 2. Superstructure optimization information flow diagram

Results

MINLP problem in Eq. 1 has been solved using MATLAB R2021a and DWSIM v6 on i5 9400 CPU 8 GB RAM Windows 10 x64 PC. In our case, the commercial methanol feedstock route results in an unprofitable process due to high purchase costs. The non-zero results of the solution are presented in Table 1.

In our case, syngas is selected as the primary raw material thanks to its low cost despite subsequent unit operations and associated operating costs. The resulting architecture proposes the highest conversion reactor for methanol production to favor the DME production and methanol which can be sold as a side product in our MINLP formulation.

Table 1. Non-zero process variables and economic measures of optimal results of Eq. 1

Plant variables									Economic Measures [M$]		
m_2 [kg/s]	m_4 [kg/s]	m_{10} [kg/s]	$m_{R2,in}$ [kg/s]	$m_{F102,B}$ [kg/s]	$m_{T101,B}$ [kg/s]	T_1 [K]	T_2 [K]	P_1 [Bar]	IC	OC	S
10.5	6.4	3.2	6.4	6.4	3.2	301	380	50	92	16	86

Temperature and pressure of flash unit maximize methanol yield while removing unreacted syngas. Produced methanol fed to medium conversion dehydration reactor to balance the DME conversion and related IC and OC. Sequential two 15-stage columns ensure the removal of high latent heat components while benefiting from the installation of large and single distillation column operations with high costs. The temperature of the feed stream as well as the temperature of the reboiler and condenser affect the efficiency of the separation process. However, feed stream temperature manipulation allows us to tune DME recovery column performance and operational costs. Overall process operation is performed at a maximum raw material rate within the defined operating window.

Conclusion

Optimum process design and synthesis under economic and environmental consideration is one of the contemporary challenges of the process systems engineering field. The idea of combining the advantages of process flowsheet simulators to obtain predictions without undergoing a significant modelling approach is useful once it is used in the black-box optimization algorithms which do not need first principal expressions explicitly. Integration of open source DWSIM simulation environment with black-box optimization solver enables us to simultaneously synthesis of process and optimization of conditions, unlike most studies which focus on commercially well-known and widely used software packages on a popular process for the architecture and processing condition calculation which accounts for feedstock selection, unit operation selection, catalysis selection, and related operating regimes. Such

a sophisticated integration of high number of process equipment alternatives and feedstocks results in a nonconvex and significantly nonlinear mixed-integer nonlinear optimization problem. As a result, the global optimality of the proposed structure is beyond the scope.

A major decision variable on the plant profitability is determined by the feedstock, which is economically syngas over a wide price spectrum under current economic specifications. Furthermore, the proposed process flow diagram is comprehensive as it enables the selection of different-size unit operations to address the capacity, installation, and operating costs to deliver a realistic evaluation to some extent. Note that, the actual process is more complex than Fig. 1 and requires a more advanced MINLP formulation which also takes uncertainties of plant variables and time-varying prices into consideration in addition to other economic considerations such as inflation and other realistic details including depreciation, land prices, and many other issues. A more detailed process synthesis and the including the impact of the price uncertainties which are characterized by logistic issues and market demands is beyond the scope. Thus, the simulations are performed for a particular and acceptable price value under marketable feed and product compositions. The architecture is flexible to address those considerations and the impact of those variables on the ultimate plant design and operations conditions can also be calculated using the proposed scheme.

$$\min_{\mathbf{m},\mathbf{y},T,P} \quad \mathrm{IC} + (\mathrm{OC} - \mathrm{S} \cdot (1 - \mathrm{f})) \cdot t_f$$

References

1. Edgar, T.F., Himmelblau, D.M., Lasdon, L.S., 2001. Optimization of Chemical Processes, 2nd ed. ed, McGraw-Hill, New York.
2. Chen, Q., Grossmann, I.E., 2017. Recent Developments and Challenges in Optimization-Based Process Synthesis. Annu. Rev. Chem. Biomol. Eng. 8, 249–283.
3. Mencarelli, L., Chen, Q., Pagot, A., Grossmann, I.E., 2020. A review on superstructure optimization approaches in process system engineering. Comput. Chem. Eng. 136, 106808.
4. Ploskas, N., Sahinidis, N. V., 2021. Review and comparison of algorithms and software for mixed-integer derivative-free optimization. J. Glob. Optim.
5. Tangsriwong, K., Lapchit, P., Kittijungjit, T., Klamrassamee, T., Sukjai, Y., Laoonual, Y., 2020. Modeling of chemical processes using commercial and open-source software: A comparison between Aspen Plus and DWSIM. IOP Conf. Ser. Earth Environ. Sci. 463, 012057.
6. Laiwatthanaphaisarn, T., Anantpinijwatna, A., 2018. Optimization of dimethyl ether production process synthesis using superstructure analysis. MATEC Web Conf. 192, 03018.
7. Saebea, D., Authayanun, S., Arpornwichanop, A., 2019. Process simulation of bio-dimethyl ether synthesis from tri-reforming of biogas: CO2 utilization. Energy 175, 36–45.
8. Nestler, F., Schütze, A.R., Ouda, M., Hadrich, M.J., Schaadt, A., Bajohr, S., Kolb, T., 2020. Kinetic modelling of methanol synthesis over commercial catalysts: A critical assessment. Chem. Eng. J. 394, 124881.
9. Bailie, R.C., Whiting, W.B., Shaeiwitz, J.A., Turton, R., Bhattacharyya, D., 2018. Analysis, Synthesis, and Design of Chemical Processes, 5th editio. ed, Prentice Hall, Boston.

5
A Simultaneous Training and Input Selection Algorithm for Classification Problems Using Piecewise Approximations

Hasan Sildir[1], Sahin Sarrafi[2], Taskin Kavzoglu[3]

[1]*Chemical Engineering, Gebze Technical University, 41400, Kocaeli, Türkiye*
[2]*SOCAR Türkiye R&D and Innovation Co., 35800, Izmir, Türkiye*
[3]*Geomatics Engineering, Gebze Technical University, 41400, Kocaeli, Türkiye*

Conferance: The 9th International Conference on the Foundations of Systems Biology in Engineering, Boston
Journal: IFAC-PapersOnLine, **2022** 55,(23), 7-12
https://www.sciencedirect.com/science/article/abs/pii/S2405896323000058
DOI: https://doi.org/10.1016/j.ifacol.2023.01.005

Abstract

Heart disease diagnosis using few measurements is a challenging an important task considering the increasing population. Artificial Neural Networks (ANNs) are promising mathematical architectures once the training is performed in an elegant manner to avoid theoretical challenges related to high nonlinearity, nonconvexity using few input variables to ensure generalization capability. This study shows the impact of the piecewise linear approximation of nonlinear functions in ANN architecture and training problem to benefit from the mixed integer linear problem formulation for the simultaneous input selection and training to obtain mixed integer programming based ANN (MIP-ANN). Proposed formulation is further tailored through linking constraints to remove the connections from the eliminated inputs to favor parameter identifiability. A publicly available dataset is considered as a case study of whose results are also compared to traditional ANN with all inputs (FC-ANN) and a relatively more straightforward but common input selection method (SKB-ANN). The results provide a comparable performance despite significant reduction in the input space in addition to significant computational and theoretical advantages thanks to advanced formulation.

Keywords: Piecewise Linear Approximation; Input Selection; Artificial Neural Networks; Classification; Mixed Integer Linear/Nonlinear Programming.

Introduction

An artificial neural network (ANN) model is formed by artificial neurons on several layers that emulate biological neurons and the synaptic connections. They are effective in identifying patterns and other underlying relationships in multidimensional data. They are also good at dealing with a large set of variables possessing non-linearity and complex structures. Each processing

node in the network receives and sums a set of input values and passes this sum through an activation function providing the output value of the node, which in turn forms one of the inputs to a processing node in the next layer. Activation functions are employed in order to decrease the number of iterations while they introduce non-linearity into the network and thus improve its performance. They can be categorized as trainable, learnable or adaptable activation functions (Apicella *et al.*, 2021).

The power of a neural network depends on how well it generalizes from the training data. The number of training samples, network architecture, number of nodes, activation function are the main factors affecting generalization capability (Kavzoglu, 1999). Pruning and input selection (or feature selection) are of considerable importance to avoid curse of dimensionality, particularly when scarcity of training samples is the case (Kavzoglu and Mather, 1998, 2000).

The determination of underlying causes and risk factors of heart diseases is a complicated task with different time and spatially dependent contributions (Turaman, 2022). With high number of patients and time consuming health tests, the monitoring and evaluation of the resulting big data become a hindering process. The efforts to diagnose the heart failure through computational tools mostly focus on data driven methods due to limited knowledge on explicit first principle expressions. The recent trends include advancements in both data and algorithm level (Ahsan and Siddique, 2022). Machine learning classifiers, after entropy based feature engineering, have been used and delivered binary classification for the heart disease (Rajkamal and Karthi, 2022). The impact of some variables for the prediction of heart disease is investigated in (Pan *et al.*, 2022). Recent trends in the prediction of heart disease are summarized in (Diwakar *et al.*, 2020).

The main purpose of the study is to investigate the combinatorial effect of training sample size, activation function, and input selection through development of piece-wise linear equations in the training problem for classification purposes. To achieve this purpose the heart disease dataset from Statlog (270 samples) including 13 attributes were employed, and accuracy assessment was then conducted.

Methodology

Artificial Neural Networks

A typical fully connected feedforward ANN expression is given by:

$$p = f_{OL}(w_{OL} f_{HL}(w_{HL} u + b_{HL}) + b_{OL}) \tag{1}$$

where f_{OL} and f_{HL} are output and hidden layer activation functions, respectively; w_{OL} and w_{HL} are output and hidden layer parameters, respectively; b_{OL} and b_{HL} are output and hidden layer bias vectors, respectively; u is the vector of inputs; p is the output vector which defines the individual probabilities when classification is under consideration. f_{OL} is limited to some kind of normalization function (e.g., softmax), unlike f_{HL} which has a lot more flexible nature. Both activation functions contribute to the nonlinearity and non-convexity of the training problem with cross-entropy error formulation in the objective function:

$$\begin{aligned} &Min_{w_{OL}, w_{HL}, b_{OL}, b_{HL}} \quad \sum_{i=1}^{N} \sum_{j=1}^{M} -y_{ij} \cdot ln(p_{ij}) \\ &s.t \\ &p_i = f_{OL}(w_{OL} f_{HL}(w_{HL} u_i + b_{HL}) + b_{OL}) \end{aligned} \tag{2}$$

where N is the number of training data; M is the number of outputs; p_i is the probability prediction vector and calculated using the i^{th} input vector, u_i. Note that f_{OL} performs the nonlinear mapping of hidden layer output, p_i^{HL}, for the normalization to obtain the individual probability of j^{th} output for the i^{th} training sample using:

$$p_{ij} = \frac{e^{p_{ij}^{HL}}}{\sum_{j=1}^{M} e^{p_{ij}^{HL}}} \tag{3}$$

where p_{ij}^{HL} is the value of the j^{th} output at i^{th} training sample before normalization by output layer activation function. The traditional cross-entropy formulation in equation (2) is further modified to yield the following nonlinear optimization problem:

$$\begin{aligned} &Min_{w_{OL}, w_{HL}, b_{OL}, b_{HL}} \quad \sum_{i=1}^{N} \sum_{j=1}^{M} -y_{ij} \cdot \left\{ p_{ij}^{HL} - ln\left(\sum_{j=1}^{M} e^{p_{ij}^{HL}} \right) \right\} \\ &s.t \\ &p_i^{HL} = w_{OL} f_{HL}(w_{HL} u_i + b_{HL}) + b_{OL} \end{aligned} \tag{4}$$

Note that equation (4) is still nonconvex regardless of the hidden layer activation function due to natural logarithm and exponential function combination in the objective function. Furthermore, some common activation functions (e.g., *tanh*) introduce additional complexity to the optimization problem and the resulting problem suffers from aforementioned issues. In addition, such problem does not address the input selection.

Piecewise-approximated ANN training problem

The modified training formulation in equation (4) contains nonlinearities in natural logarithm, exponential term in addition to activation function all of which require a piecewise approximation to ensure a mixed integer linear programming (MILP) formulation. This study focuses on hyperbolic tangent and rectified linear unit activation function, which have significant applications in machine learning. However, the proposed formulation is easily adaptable to other functions, as well. Fig. 1 shows the approximation of the hyperbolic tangent, PWLtanh(x), using six discretization points and traditional *tanh* function for approximation capability demonstration. Note that the piecewise representation is constrained to the region where activation function slope is non-zero in order to reduce the activation function saturation problem during the training.

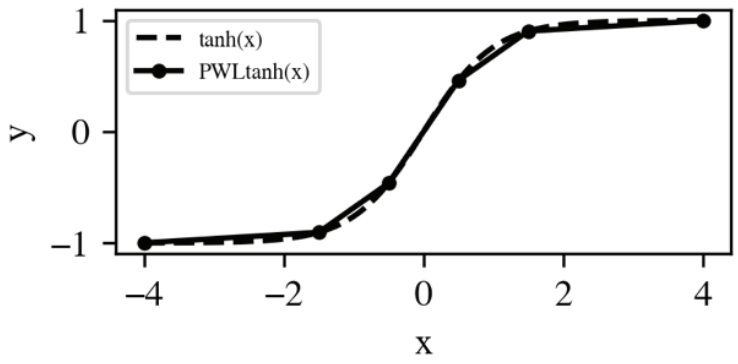

Figure 1. Piecewise linear approximation of hyperbolic tangent

Evaluation of a particular point using the PWLtanh(x) is calculated from the solution of MILP in (5):

$$
\begin{gathered}
Min_{\lambda,\beta} \quad z \\
s.t \\
-4\lambda_1 - 1.5\lambda_2 - 0.5\lambda_3 + 0.5\lambda_4 + 1.5\lambda_5 + 4\lambda_6 = q \\
-0.99\lambda_1 - 0.9\lambda_2 - 0.46\lambda_3 + 0.46\lambda_4 + 0.9\lambda_5 + 0.99\lambda_6 = z \\
\sum_{k=1}^{6} \lambda_k = 1 \\
\lambda_1 \leq \beta_1 \\
\lambda_2 \leq \beta_1 + \beta_2 \\
\lambda_3 \leq \beta_2 + \beta_3 \\
\lambda_4 \leq \beta_3 + \beta_4 \\
\lambda_5 \leq \beta_4 + \beta_5 \\
\lambda_6 \leq \beta_5 \\
0 \leq \lambda_k, k = 1,\ldots,6 \\
\beta_m \epsilon \{0,1\}, m = 1,\ldots,5
\end{gathered}
\tag{5}
$$

where q is the value at which the hyperbolic tangent evaluation is desired; z is the result of the evaluation of PWLtanh(x); λ and β are additionally defined dummy variables to include a combination of piecewise lines based on the evaluation point and to ensure the convexity. The feasibility problem in (5) would deliver a better approximation of the original tanh function using more discretization points but brings about significant computational load due to the introduction of more variables.

Unlike *tanh*, *ReLU* which already has a piecewise linear shape, can be represented perfectly using the formulation in equation (6) with relatively fewer number of additional variables. In addition, the discontinuity at the origin, which causes some problems in the rigorous nonlinear optimization solutions, is also considered and handled in the proposed MILP formulation.

$$
\begin{gathered}
Min_{\lambda,\beta} \quad z \\
s.t \\
-4\lambda_1 + 4\lambda_3 = q \\
4\lambda_3 = z \\
\lambda_1 + \lambda_2 + \lambda_3 = 1 \\
\lambda_1 \leq \beta_1 \\
\lambda_2 \leq \beta_1 + \beta_2 \\
\lambda_3 \leq \beta_2 \\
0 \leq \lambda_k, k = 1,2,3 \\
\beta_m \epsilon \{0,1\}, m = 1,2
\end{gathered}
\tag{6}
$$

A similar formulation is derived for exponential function, although it is a convex function itself, for the MILP formulation, using the breakpoints given in Fig. 2 which shows the PWLexp(x), a piece-wise linear approximation to exponential function.

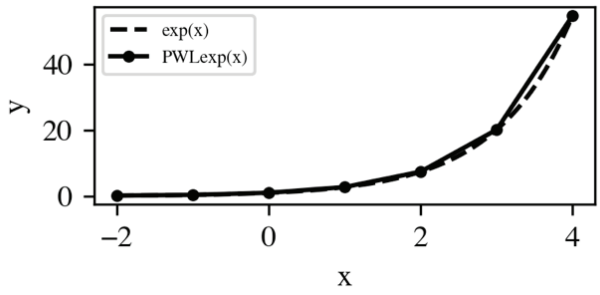

Figure 2. Piecewise linear approximation of exponential function

Similarly, the approximation of logarithm function, PWLlog(x), is shown in Fig. 3.

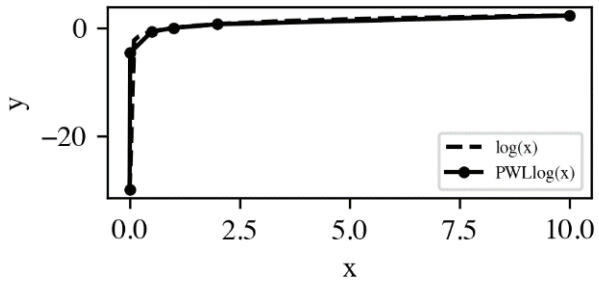

Figure 3. Piecewise linear approximation of logarithm function

Simultaneous ANN training and input selection formula

Equation (4), with some modification in the objective function compared to the traditional cross-entropy formulation, does not address the input selection issue and local solutions during the training. Equation (7) contains all aforementioned piece-wise linear approximation expressions for nonlinear and nonconvex functions in addition to linking constraints to eliminate the impact of an input once it is not selected during the optimization.

$$
\begin{gathered}
Min_{w_{HL},b_{OL},b_{HL},u_b} \quad F \\
s.t \\
F = \sum_{i=1}^{N}\sum_{j=1}^{M} -y_{ij} \cdot \left\{ p_{ij}^{HL} - PWLlog\left(\sum_{j=1}^{M} PWLexp(p_{ij}^{HL}) \right) \right\} \\
p_i^{HL} = w_{OL} PWLf_{HL}(w_{HL} u_i + b_{HL}) + b_{OL} \\
w_{HL,ij} \leq u_{b,j} \cdot w_{HL,max}, i = 1,\ldots,K, j = 1,\ldots,R \\
u_{b,j} \cdot w_{HL,min} \leq w_{HL,ij}, i = 1,\ldots,K, j = 1,\ldots,R \\
\sum_i u_{b,i} = u_{desired} \\
u_{b,i} \in \{0,1\}, i = 1,\ldots,R
\end{gathered}
\tag{7}
$$

where u_b is the binary variable to represent the existence of an input; $u_{desired}$ is the number of inputs to be included in the training problem; $w_{HL,min}$ and $w_{HL,max}$ are the lower and upper bounds of w_{HL}, respectively; K is the number of hidden layer neurons; R is the number of available inputs; $PWLf_{HL}$ is the piecewise approximation of the activation function. Note that the binary variable for the input selection, u_b, tightens the optimization search space through constraining the lower and upper bounds of the corresponding column of the w_{HL} to zero, eliminating the information flow from particular input to the hidden layer, although it is not included in the ANN evaluation explicitly. Obtaining such a result is challenging for traditional regularization techniques as they do focus on the penalization of the weight magnitudes rather than eliminating them. Finally, note that w_{OL} is not a decision variable in (7) to ensure MILP formulation by elimination of continuous variable multiplications and obtained after a fully connected ANN training, similar to idea proposed by extreme machine learning approach (Akyol, 2020).

Formulation in (7) enables the solution of the optimization problem using MILP algorithms benefiting from a wide range of decomposition approaches and piecewise convex nature during the iterations while performing relaxations. Such a superiority to deal with sophisticated and complicated equations in the classification problems enable the implementation of the feasibility cuts to tighten the search region during the iterations and ensure the global optimality once the problem is solved to satisfactory optimality gap (Wittmann-Hohlbein and Pistikopoulos, 2013). Once related equations are represented by enough breakpoints during the piecewise approximation, the corresponding optimum is also a satisfactory solution to original nonconvex mixed integer nonlinear problem (MINLP), whose global solution, especially at large-scale, is a challenging and computationally costly task. On the other hand, the proposed MILP formulation is strongly desirable, besides global optimality considerations, due to advanced decomposition techniques and computational capabilities, including parallelization, which are already implemented in solvers (Anand, Aggarwal and Kumar, 2017). In addition, the method is also useful to handle discontinuities in the activation function (e.g., ReLU), which might introduce additional problems when NLP solvers are used for traditional training problem.

Table 1. Selected inputs for different training ratios, activation functions and architectures

	ReLU				***tanh***			
	SKB**-ANN**		**MIP-ANN**		SKB **-ANN**		**MIP-ANN**	
Training Sample Ratio [%]	**10**	**50**	**10**	**50**	**10**	**50**	**10**	**50**
Age	■							
Sex								
chest pain type		■		■		■	■	■
resting blood pressure								
serum cholesterol in mg/dl								
fasting blood sugar > 120 mg/dl								
resting electrocardiographic results								
maximum heart rate achieved		■						
exercise induced angina	■		■			■		
oldpeak								
the slope of the peak exercise					■			
number of major vessels								
thal			■	■	■		■	■

A major advantage of the proposed formulation is the simultaneous training and feature selection capability. Thus, a subsequent and decentralized input selection and training procedures are not necessary unlike the traditional sequential approach in which the training is performed after the inputs are selected using a relatively simpler method. Such a sequential approach might deliver poor test performance due to theoretical discrepancies between input selection mathematical approach and prediction model. Thus, the proposed method is superior as the inputs are selected taking the ANN architecture to be used for test purposes into account. In addition, the method provides additional robustness in terms of parameter identifiability issues, especially once the training sample size is small, by eliminating the connections from the input neurons to hidden layer. Although (7) does not introduce additional pruning capability yet, it can easily be extended for ANN architecture compression purposes through addition of more binary variables which represent the existence of connections in addition to already proposed binary variables.

Result

In this study, a novel methodology employing input selection and piece-wise approximation strategies were employed to obtain robust network solutions on publicly available and widely used classification data (Yazid *et al.*, 2018). The main objectives of the research conducted here are threefold. Performances of SKB-ANN, which employs an easy-to-use data driven highest scoring input selection method based on the mutual info (Kozachenko and Leonenko, 1987),

and MIP-ANN architecture, which is obtained from using the mixed integer problem in (7) were compared to that of fully connected network with all inputs. The hypothetical challenge for the comparison includes the selection of very few number of inputs to exploit and demonstrate the potential of the proposed methodology in addition to address real-time management problems in biological systems where high number of measurements is cumbersome. In addition to the analysis of different architectures, the effect of activation functions (*ReLU* and *tanh*) was also analyzed for two training sample propositions. For representing limited data case, %10 of the samples was employed at the training stage. For the other case, half of the sample was employed for training. Equation (7) was solved using Gurobi (Gurobi Optimization, 2018) solver to which communication is performed through PYOMO (Hart, Watson and Woodruff, 2011).

To assess the robustness of the input selection methodology adapted in this study SKB-ANN and MIP-ANN architectures with different sampling rates and activation functions were set to search the two inputs that best represent the whole dataset (Table 1). Out of 13 input features, two features of the dataset were selected both algorithms. When the frequency of the selected inputs was considered, it was seen that chest pain type and thal were the most commonly (six times) were selected. This clearly show their high representativeness compared with the other features. Exercise induced angina, age, maximum heart rate achieved, the slope of the peak exercise are the other selected features. It is interesting to see that sex, resting blood pressure, and cholesterol level were not selected in any case. With the use of 50% of the whole dataset, chest pain type was always selected, which shows its importance in heart attack cases. Compared with SKB-ANN, proposed MIP-ANN was more robust in selecting input features based on training ratio and activation function changes.

Combinatory effect of network type (i.e., FC-ANN, SKB-ANN, MIP-ANN), activation function (*ReLU* and *tanh*) and sample size (10% and 50% ratios) on the classification performance were assessed on both training and testing datasets (Table 2). Although the highest classification accuracies were achieved by the use of fully connected network with the use of *tanh* function (»95% accuracy for training samples), the same setting yielded the lowest accuracies for the test data, which can be described as a typical overfitting behavior for large networks. Comparison of SKB-ANN and proposed MIP-ANN performances revealed that an accuracy improvement of 5-11% was obvious for the test dataset. It should be also mentioned that the use of *tanh* function outperformed the *ReLU* function with slight difference. Finally, it is interesting to see that the MIP-ANN architecture performed better with 10% of the dataset, which shows its superiority on limited training datasets.

Table 2. Classification performance comparison for FC-ANN, SKB-ANN and MIP-ANN using only two selected inputs at 10% and 50% training ratios (TR)

				FC-ANN			SKB-ANN			MIP-ANN		
				C_1	C_2	Acc.	C_1	C_2	Acc.	C_1	C_2	Acc.
ReLU	10% TR.	Tr	C_1	14	1	0.889	14	1	0.815	14	1	0.778
			C_2	2	10		4	8		5	7	
		Ts	C_1	99	36	0.662	112	23	0.654	105	30	0.761
			C_2	46	62		61	47		28	80	
	50% TR.	Tr	C_1	63	6	0.857	58	11	0.785	62	7	0.793
			C_2	9	57		18	48		21	45	
		Ts	C_1	71	10	0.800	54	27	0.681	67	14	0.741
			C_2	17	37		16	38		21	33	
tanh	10% TR.	Tr	C_1	14	1	0.963	11	4	0.778	14	1	0.778
			C_2	0	12		2	10		5	7	
		Ts	C_1	89	46	0.617	71	64	0.694	112	23	0.778
			C_2	47	61		10	98		31	77	
	50% TR.	Tr	C_1	65	4	0.948	54	15	0.800	62	7	0.793
			C_2	3	63		12	54		21	45	
		Ts	C_1	63	18	0.756	58	23	0.704	67	14	0.741
			C_2	15	39		17	37		21	33	

Conclusions

Data driven empirical models, since the inherent expressions do not specifically rely on the actual nature of the phenomena, are useful only once the training is performed after addressing several theoretical and practical challenges. Feedforward ANNs have a flexible mathematical capability to represent complex interactions among various data sources, including the systems biology field. Such flexibility is obtained through sophisticated integration of information from the inputs through nonlinear and nonconvex activation functions, resulting in high number of local solutions in the training problem and causing reproducibility issues. The proposed formulation still benefits from the nonlinear properties of the ANN expressions but eliminates the nonconvexity problems through a piecewise linearization approach (Sildir and

Aydin, 2022). In theory, once the number of breakpoints is high for a particular nonlinear function, a satisfactorily similar performance is obtained with significant advantages in the related optimization problem. This study focuses on the development of such an approach for the classification problems and demonstrates the capability through publicly available dataset.

To the best of authors' knowledge, the paper is novel in terms of addressing the simultaneous training and input selection on classification problems through mathematical reformulations in the traditional ANN training problem combined with high number of piecewise representations of nonlinear and nonconvex functions. Furthermore, the resulting MILP is a sophisticated alternative to traditional methods which employ a sequential approach with different mathematical considerations in the input selection and data driven model training, lacking the theoretical compatibility guarantee between the two. In our cases, despite different initial guesses for the ANN parameters and inputs, the converged architecture and weights are same and test performance is similar to training thanks to simultaneous training and input selection based on the actual ANN architecture. In addition, a computationally scalable and parallelizable optimization problem becomes available due to advanced decomposition and reformulation methods which are already embedded in existing software. This is an important aspect for the application to larger dataset and ANN architecture since the approach requires additional binary and continuous variable for the piecewise representation of nonlinear expressions. In our case, several breakpoints have resulted in satisfactory accuracy for approximation and can easily be increased once a better representation of the nonlinearity is required.

The rigorous formulation in the optimization problem, which includes the existence of inputs through binary variables and related linking constraints, delivered a more robust input selection performance at different training ratios or activation functions, compared to a common input selection algorithm which employs a statistical criterion on the data directly without evaluating the impact of the ANN architecture inherently. In addition, the method shows a better impact when low number of training samples is available. However, in all cases in this study, the training and the test performances are similar once the proposed formulation is used under different training ratios and activation functions. In other words, the method delivers an ANN architecture which does not suffer from overfitting problem, in our case, and better generalization capability. Note that such generalization does not require an explicit weight regularization and obtained through linking constraints in the training problem, but formulation is flexible to further penalization of the ANN weights, if desired.

Besides mathematical contributions, the study focuses on the very sparse representation of the available inputs and theoretical highest predictive

potential of the dataset is not the ultimate objective. However, it is a future work and should cover additional considerations on number of hidden neurons and layers in addition to data pre-processing at different training samples considering advanced shuffling techniques. This study is limited to hypothetical case in which a small architecture, which contains few neurons and, a practical and useful approach when high number of samples are employed in real-time such as heart disease diagnosis at a large population.

References

Ahsan, M. M. and Siddique, Z. (2022) 'Machine learning-based heart disease diagnosis: A systematic literature review', *Artificial Intelligence in Medicine*, 128, p. 102289. doi: https://doi.org/10.1016/j.artmed.2022.102289.

Akyol, K. (2020) 'Comparing of deep neural networks and extreme learning machines based on growing and pruning approach', *Expert Systems with Applications*, 140. doi: 10.1016/j.eswa.2019.112875.

Anand, R., Aggarwal, D. and Kumar, V. (2017) 'A comparative analysis of optimization solvers', *Journal of Statistics and Management Systems*, 20(4), pp. 623–635.

Apicella, A. *et al.* (2021) 'A survey on modern trainable activation functions', *Neural Networks*, 138, pp. 14–32.

Diwakar, M. *et al.* (2020) 'Latest trends on heart disease prediction using machine learning and image fusion', *Materials Today: Proceedings*, 37(Part 2), pp. 3213–3218. doi: 10.1016/j.matpr.2020.09.078.

Gurobi Optimization, L. L. C. (2018) 'Gurobi optimizer reference manual'.

Hart, W. E., Watson, J. P. and Woodruff, D. L. (2011) 'Pyomo: Modeling and solving mathematical programs in Python', *Mathematical Programming Computation*, 3(3), pp. 219–260. doi: 10.1007/s12532-011-0026-8.

Kavzoglu, T. (1999) 'Determining optimum structure for artificial neural networks', in *Proceedings of the 25th Annual Technical Conference and Exhibition of the Remote Sensing Society*. Citeseer, pp. 675–682.

Kavzoglu, T. and Mather, P. M. (1998) 'Assessing artificial neural network pruning algorithms', in *Proceedings of the 24th Annual Conference and Exhibition of the Remote Sensing Society*. Greenwich, UK, pp. 603–609.

Kavzoglu, T. and Mather, P. M. (2000) 'Using feature selection techniques to produce smaller neural networks with better generalisation capabilities', in *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*. IEEE, pp. 3069–3071.

Kozachenko, L. F. and Leonenko, N. N. (1987) 'Sample estimate of the entropy of a random vector', *Problemy Peredachi Informatsii*, 23(2), pp. 9–16.

Pan, C. *et al.* (2022) 'Impact of categorical and numerical features in ensemble machine learning frameworks for heart disease prediction', *Biomedical Signal Processing and Control*, 76, p. 103666. doi: 10.1016/j.bspc.2022.103666.

Rajkamal, R. and Karthi, A. (2022) 'Heart Disease Prediction Using Entropy Based Feature Engineering and Ensembling of Machine Learning Classifiers', *Expert Systems with Applications*, p. 117882. doi: https://doi.org/10.1016/j.eswa.2022.117882.

Sildir, H. and Aydin, E. (2022) 'A Mixed-Integer linear programming based training and feature selection method for artificial neural networks using piece-wise linear approximations', *Chemical Engineering Science*, 249, p. 117273. doi: 10.1016/j.ces.2021.117273.

Turaman, C. (2022) 'Classification of the risk factors of coronary heart disease and their evolutionary origins', *Health Sciences Review*, 3, p. 100027. doi: 10.1016/j.hsr.2022.100027.

Wittmann-Hohlbein, M. and Pistikopoulos, E. N. (2013) 'On the global solution of multi-parametric mixed integer linear programming problems', *Journal of Global Optimization*, 57(1), pp. 51–73.

Yazid, M. H. A. *et al.* (2018) 'Artificial neural network parameter tuning framework for heart disease classification', in *2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE, pp. 674–679.

6
Architectural Design of Chemical and Biological Pathways through Parameter Sensitivity Oriented Mixed Integer Formulations

Emir Topac[a], Emrullah Erturk[a], Ozgun Deliismail[b], Sahin Sarrafi[b], Hasan Sildir[a]

[a]*Department of Chemical Engineering, Gebze Technical University, Kocaeli, 41400, Türkiye*
[b]*SOCAR Türkiye R&D and Innovation Co., Izmir, 35800, Türkiye*
hasansildir@gtu.edu.tr

Conferance: **33rd European Symposium on Computer Aided Process Engineering (ESCAPE33), Athens, Greece**

Journal:

DOI:

Abstract

A mixed integer nonlinear programming (MINLP) formulation is developed for the automatized design of chemical and biological pathways. The proposed formulation is solved using rigorous algorithms and enables the selection of reactions whose rate constants have the highest sensitivity. Unlike the traditional parameter estimation approach, the training performance is introduced as constraint which is pre-defined to balance the architectural simplicity and fitting performance. Adjoint sensitivity expressions are additionally solved simultaneously in the formulation and the resulting parameter sensitivities are included in the objective. Binary variables are introduced in the linking constraints for the selection of a reaction subset from a more complex pathway, also tightening the search space for the optimization algorithm. The approach is implemented on a complex chemical/biological pathway which processes glycose as a primary feedstock. The ultimate and reduced pathway prediction bounds are significantly narrower and deliver similar mean predictions compared to large and full pathway, despite 25% reduction in available reaction paths in the superstructure.

Keywords: Parameter identifiability; Mixed integer programming; Sensitivity; Uncertainty.

Introduction

Chemical and biological pathway models include a high number of rate expressions with tunable parameters to be estimated from the experimental data. The lack of spatial and temporal measurements might result in statistically ill-defined inverse problems with multiple solutions in parameter

estimation (Mclean and Mcauley, 2012). Furthermore, the computational load and identifiability problems increase once the disturbances and measurement noises appear especially when the model architecture is large.

Usually, trial and error procedures or sequential methods are applied for the architectural synthesis and parameter identifiability tasks, which require significant manual effort in addition to decentralized computations on theoretical level. In (Ramadan et al., 2018) the most sensitive parameter selection utilizing LASSO and applied to a biomechanical system. Shapiro et *al.* developed a heuristic algorithm to eliminate some parameters and re-calculate the sensitivities until desired condition value is obtained (Shapiro et al., 2014).

A simultaneous approach using rigorous formulations to obtain the pathway architecture and corresponding parameters based on the parameter sensitivity values and their contribution to uncertainty is scarce, and would have the potential to exploit available statistical knowledge from experimental data for a better model development. This study focuses on the development of a mixed-integer formulation to perform architectural synthesis and parameter estimation tasks based on sensitivities in addition to the model training performance.

Methodology

A typical reaction pathway is represented by ODEs and given by:

$$\frac{dy}{dt} = f(t, y, p) \tag{1}$$

where y is the vector of outputs which are typically concentrations; p is the vector of time-invariant parameters; t is the time; f is the vector of mathematical rate expressions.

The derivative of the terms in Eq. 1 results in forward sensitivity ODE for a particular parameter and output:

$$\frac{ds_{ij}}{dt} = \frac{df_j}{dy_i} s_{ij} + \frac{df_j}{dp_i} \tag{2}$$

where s_{ij} is the sensitivity of j^{th} output, y_j, to i^{th} parameter, p_i. A simultaneous solution of Eq. 1 and Eq. 2 is required to address the specific considerations on parameter selection process to account for the model performance, additionally.

A modified mixed integer formulation is developed for the architectural design and estimation of its parameters who have the highest sensitivity, eliminating the parameters with relatively smaller impact on the outputs. The formulation is given by:

$$
\begin{aligned}
& Min_{p,b} \quad -\sum s_{ij,k}^2 \\
& s.t. \\
& \frac{dy}{dt} = f(t, c, p) \\
& \frac{ds_{ij}}{dt} = \frac{df_j}{dy_j} s_{ij} + \frac{df_j}{dp_i} \quad i = 1,2, \dots, M \; j = 1,2, \dots, P \\
& \sum \left\| y_k - y_{m,k} \right\| \leq \varepsilon \\
& b_i p_{i,min} \leq p_i \leq b_i p_{i,max} \qquad i = 1,2, \dots, P \\
& b_{min} \leq \sum b_i \leq b_{max} \quad b_i \in \{0,1\} \quad i = 1,2, \dots, P
\end{aligned}
\tag{3}
$$

where $s_{ij,k}$ is the sensitivity at sample time k; y_k is the model prediction at k; $y_{m,k}$ is the measurement at k; e is the desired maximum training error; $p_{i,min}$ and $p_{i,max}$ are lower and upper bounds of i^{th} parameter, p_i, respectively; b_i is the binary variable to account for the existence of p_i; b_{min} and b_{max} are lower and upper bounds of number of parameters desired in the ultimate architecture.

Solution of Eq. 3 requires an explicit formulation for sensitivity ODEs for the rigorous approach, which calls for an automatic or symbolic differentiation. There are various solver packages which deliver such expressions (Andersson et al., 2019), which in turn mathematical reformulations and additional error from numerical approximations are avoided. On the other hand, sensitivity equations are mostly nonlinear and further mathematical reformulations and decompositions, in addition to dense discretization, might be essential for pathways with high complexity. In contrast to nonlinear optimization which applies to parameter estimation with fixed architectures, the proposed mixed integer formulation includes binary variables to account for the existence of particular parameter values which represent the reaction rate constant once they have a nonzero value. Linking constraints are also introduced to tighten the search space by fixing the parameter values while eliminating the corresponding reaction path from the network. The formulation is flexible and can address measurements and sensitivity considerations at desired sample times. Current formulation includes the selection of the parameters which deliver the maximum squared sensitivity at measurement points.

The parameter covariance matrix for a particular architecture (c_p) is subsequently calculated using bootstrapping method (Godo-Pla et al., 2019) for further analysis on the identifiability issues. The uncertainty in the outputs due to parameter variations is calculated using (Tellinghuisen, 2001):

$$c_y = S_k c_p S_k^T \tag{4}$$

where c_y is the output covariance matrix; S_k is the sensitivity matrix at k.

The approach is implemented on a complex chemical/biological pathway and solved using BONMIN solver using PYOMO computation environment (Hart et al., 2011).

Results

The proposed approach is implemented on 5-Hydroxymethylfurfural (HMF) synthesis from glycose (Tang et al., 2017) which can be obtained through biological processes, providing a promising opportunity for sustainability considerations. The full network superstructure is shown in Fig. 1a, where several different reaction paths enable the production of a particular compound, which in turn contributes to identifiability of these paths. With a coupled and dependent set of ODEs, the unique estimation of the parameters becomes a challenging task and the proposed formulation is implemented to demonstrate the impact, unlike (Erturk et al., 2021) where sensitivity considerations are not included. Thus, in addition to six differential equations to represent the compounds, 72 additional sensitivity ODEs are introduced to account for 12 parameters in the full network. Once the simultaneous architectural synthesis and the parameter estimation are performed based on the selection of the most sensitive parameters, the resulting reduced network is presented in Fig. 1b.

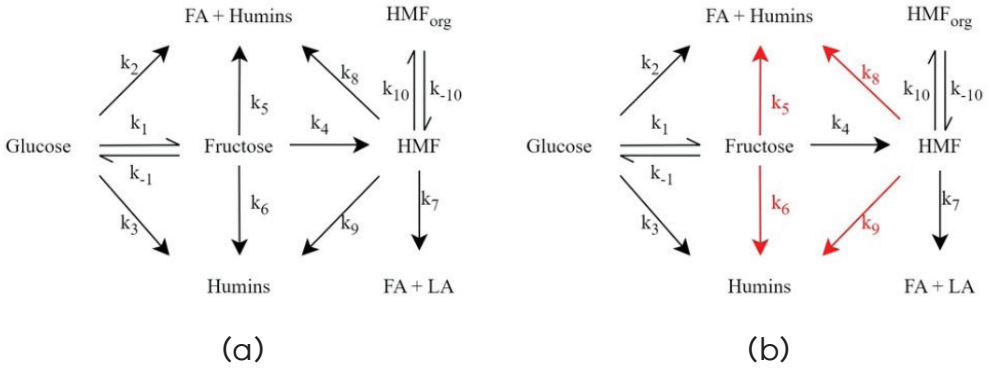

Figure 1. Reaction pathway
Full network (b) Reduced network

The eliminated reaction paths from Fig. 1a are shown in Fig. 1b using red arrows. Next, the performance of the full network and the reduced network is compared in terms of prediction similarity and uncertainty. The uncertainty for the solutions is evaluated from the diagonal elements of c_y which is calculated from Eq. 4 after bootstrapping procedures are performed.

Fig. 2a and Fig. 2b include solutions of pathway ODEs for the full network and the reduced network, respectively, for Fructose and FA+Humins. The mean predictions, which exclude the parameter covariance matrix and related propagation equations, show similar profile due to constraint in Eq. 3 to deliver a satisfactory training performance characterized by *e*. In turn, a similar profile is obtained despite elimination of four reaction paths as shown in Fig. 1b.

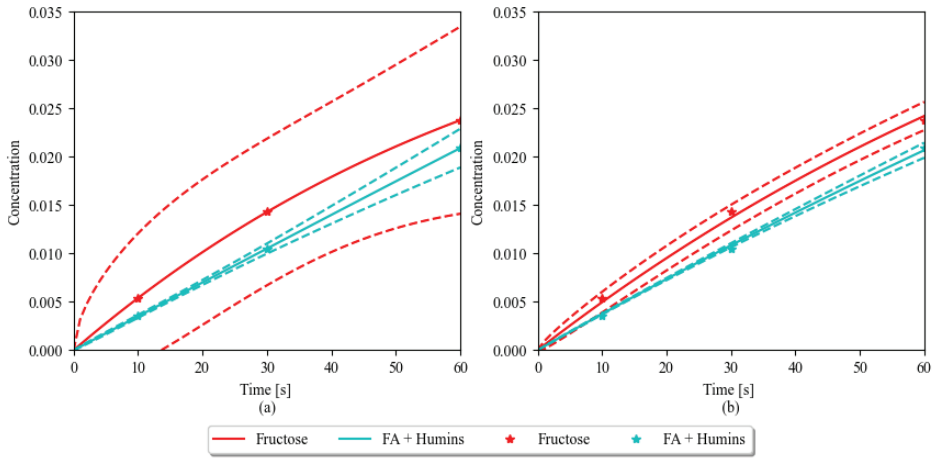

Figure 2. Mean solution and uncertainty interval for Fructose, FA+ Humins (a) Full network (b) Reduced network

As a result of numerical scale differences among pathway variables, the solution profiles for HMF_{aq} and HMF_{org}, are presented in Fig. 3b, from which the latter is for the reduced network. Compared to the other components, the impact on uncertainty on HMF_{org} is less significant as the corresponding component related reactions have not changed (see Fig. 1a and Fig. 1b). However, some decrease in the uncertainty range is still observed due to a more robust parameter estimation which accounts for the identifiability.

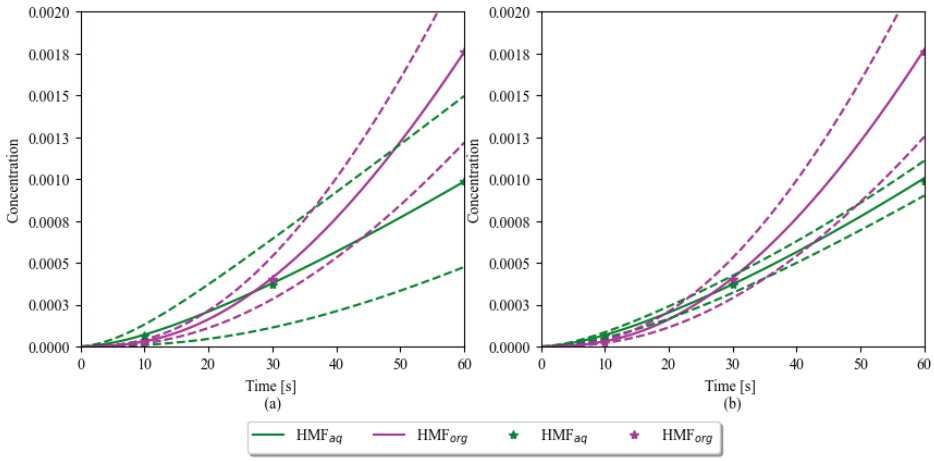

Figure 3. Mean solution and uncertainty interval for HMF_{aq}, HMF_{org} (a) Full network (b) Reduced network

Table 1 includes a comparison on the maximum and minimum dependent variable values based on uncertainty at final time. The prediction bounds for the reduced model is significantly more tightened compared to the full network. However, a more drastic reduction in uncertainty is favored by the elimination of the reaction paths associated with a particular component.

Table 1. Impact on output uncertainty

	Full Network [´100]		Reduced Network [´100]		Difference [%]
	Min	*Max*	*Min*	*Max*	
Glucose	19.15	21.16	19.96	20.32	82.09
Fructose	1.41	3.35	2.28	2.56	85.57
HMF_{aq}	0.05	0.15	0.09	0.11	80.58
HMF_{org}	0.12	0.23	0.13	0.23	5.45
FA+Humins	1.89	2.29	1.99	2.14	62.50
FA+LA	0.00	0.10	0.00	0.09	10.00

Discussion

Traditional parameter estimation problems for complex pathways usually process a fixed architecture and include the training error in the objective function under physically constrained rate expression parameters. In contrast to traditional approach, the proposed mixed-integer formulation performs simultaneous architectural design and parameter estimation by adding binary variables which consider the existence of a particular path and tighten the search space during the optimization through linking constraints. The model training performance is introduced as a constraint to balance the structural reduction and fitting performance, since elimination of some reaction paths

from the superstructure hinders the fitting capability, in theory. Moreover, the simultaneous evaluation and consideration of sensitivity expressions enable the calculations to be performed by rigorous optimization algorithms, which can include advanced mathematical reformulations, decompositions, and approximations to account for various sensitivity requirements. Such a centralized processing of the complex problem delivers the best configuration and high identifiable parameters, especially when global or sophisticated solvers are utilized.

As a downside, the approach calls for more computational load compared to traditional methods, due to binary variables to represent the existence of parameters, evaluation of higher number of equations during optimization, additional nonlinearity terms, the rigorous solution algorithm delivers a more robust model architecture. Thus, subsequent tasks related to model update or control tasks become computationally more feasible.

A biochemical pathway is considered to demonstrate the impact of the approach. However, the problem formulation is flexible and can be further tailored to various applications. Current focus includes the development of sophisticated mathematical reformulations to address the computational problems when larger networks with more parameters are needed.

References

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., Diehl, M., 2019. CasADi: a software framework for nonlinear optimization and optimal control. Math. Program. Comput. 11, 1–36.

Erturk, E., Erdal, A., Sildir, H., 2021. Tam Sayili ve Surekli Optimizasyon Problemi ile Reaksiyon Ag Modellerinin Kucultulmesi. Konya Muhendis. Bilim. Derg. 9, 142–156.

Godo-Pla, L., Emiliano, P., Valero, F., Poch, M., Sin, G., Monclús, H., 2019. Predicting the oxidant demand in full-scale drinking water treatment using an artificial neural network: Uncertainty and sensitivity analysis. Process Saf. Environ. Prot. 125, 317–327.

Hart, W.E., Watson, J.P., Woodruff, D.L., 2011. Pyomo: Modeling and solving mathematical programs in Python. Math. Program. Comput. 3, 219–260. https://doi.org/10.1007/s12532-011-0026-8

Mclean, K.A.P., Mcauley, K.B., 2012. Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. Can. J. Chem. Eng. 90, 351–366. https://doi.org/10.1002/cjce.20660

Ramadan, A., Boss, C., Choi, J., Peter Reeves, N., Cholewicki, J., Popovich, J.M., Radcliffe, C.J., 2018. Selecting sensitive parameter subsets in dynamical models with application to biomechanical system identification. J. Biomech. Eng. 140.

Shapiro, B., Barton, M., Mavalvala, N., Mittleman, R., Youcef-Toumi, K., 2014. Selection of important parameters using uncertainty and sensitivity analysis. IEEE/ASME Trans. Mechatronics 20, 13–23.

Tang, J., Zhu, L., Fu, X., Dai, J., Guo, X., Hu, C., 2017. Insights into the kinetics and reaction network of aluminum chloride-catalyzed conversion of glucose in NaCl–H2O/THF biphasic system. ACS Catal. 7, 256–266.

Tellinghuisen, J., 2001. Statistical error propagation. J. Phys. Chem. A 105, 3917–3921. https://doi.org/10.1021/jp003484u

Can Pyrolysis Oil Be Used as a Feedstock to Close the Gap in the Circular Economy of Polyolefins?

Berrak Erkmen[1], Adem Ozdogan[2], Ayhan Ezdesir[1], Gokhan Celik[3*]

[1] *SOCAR Türkiye R&D and Innovation Inc., Product Development Department, Izmir 35800, Türkiye*
[2] *SOCAR Türkiye, Refinery and Petrochemicals Business Unit, Ethylene and Aromatics Laboratory, Izmir 35800, Türkiye*
[3] *Middle East Technical University, Department of Chemical Engineering, Ankara 06800, Türkiye*
[*] *gcelik@metu.edu.tr*

Journal: Polymers **2023**, 15(4), 859.

https://www.mdpi.com/2073-4360/15/4/859

DOI: https://doi.org/10.3390/polym15040859

Abstract

Plastics are engineering marvels that have found widespread use in all aspects of modern life. However, poor waste management practices and inefficient recycling technologies, along with their extremely high durability, have caused one of the major environmental problems facing humankind: waste plastic pollution. The upcycling of waste plastics to chemical feedstock to produce virgin plastics has emerged as a viable option to mitigate the adverse effects of plastic pollution and close the gap in the circular economy of plastics. Pyrolysis is considered a chemical recycling technology to upcycle waste plastics. Yet, whether pyrolysis as a stand-alone technology can achieve true circularity or not requires further investigation. In this study, we analyzed and critically evaluated whether oil obtained from the non-catalytic pyrolysis of virgin polypropylene (PP) can be used as a feedstock for naphtha crackers to produce olefins, and subsequently polyolefins, without undermining the circular economy and resource efficiency. Two different pyrolysis oils were obtained from a pyrolysis plant and compared with light and heavy naphtha by a combination of physical and chromatographic methods, in accordance with established standards. The results demonstrate that pyrolysis oil consists of mostly cyclic olefins with a bromine number of 85 to 304, whereas light naphtha consists of mostly paraffinic hydrocarbons with a very low olefinic content and a bromine number around 1. Owing to the compositional differences, pyrolysis oil studied herein is completely different than naphtha in terms of hydrocarbon composition and cannot be used as a feedstock for commercial naphtha crackers to produce olefins. The findings are of particular importance to evaluating different chemical recycling opportunities with respect to true circularity and may serve as a benchmark to determine whether liquids obtained from different polyolefin recycling technologies are compatible with existing industrial steam crackers' feedstock.

Keywords: pyrolysis, pyrolysis oil, naphtha, waste plastic, chemical recycling, polyolefins, bromine number, PIONA, circular economy

Introduction

Plastics are ubiquitous and versatile materials, and they are used in all aspects of modern civilization at tremendous quantities. Four hundred million tons (400 Mt) of plastics are produced each year. It is estimated that the production of plastics will only increase and exceed one million tons by the end of 2050 [1–5]. After being used for their intended purposes, plastics complete their useful lifecycles and are discarded. Some plastics such as straws and utensils are produced for single-use with a useful lifetime between seconds to minutes, whereas some plastics such as shampoo bottles or garbage bins can be used for longer durations with a useful lifetime between weeks to years. Regardless of the time scale at which plastic is discarded, their waste management becomes crucial because they degrade slowly. For instance, it takes up to 200 years for a plastic straw to degrade naturally [6].

Although a simple comparison between useful lifetimes and natural degradation durations points out the importance of waste management of plastics to prevent their haphazard accumulations in the environment, existing infrastructures could not cope with the waste plastics, causing one of the biggest environmental challenges facing humankind: plastic pollution [1–3,5]. Global mass production analysis [7] of plastics shows that out of eight billion metric tons of plastics ever produced by 2017, 70% of them ended up in landfills or in aquatic life polluting our planet. 14% of them was incinerated to produce energy. Yet, this option causes emission of greenhouses gases such as CO_2 at large volumes. The remaining 16% was recycled to obtain lower-value materials with a low efficacy [8]. These methods are not environmentally friendly and cause continuous consumption of natural resources (from crude oil or natural gas to plastics to waste plastics), favoring a linear economy.

Plastics are, however, engineering marvels with high energy and chemical content. Producing 400 Mt plastics approximately require a consumption of 7% of crude oil and natural gas produced [3]. Considering the expensive and scarce fossil fuel resources used in plastic production, it is unfortunate to waste these resources for creating waste, polluting the environment, or losing their value to low-quality recycled products, which are oftentimes cannot be recycled after several life cycles [9–11]. This line of process can be considered as an example of linear economy because resources are linearly converted to plastics and eventually to waste. Although the linear system creates a massive economic value by producing and selling plastics, the result is the global waste plastic problem and loss of value and resources. New and innovative approaches are needed to replace the linear economy of plastics by the circular economy. There are currently a lot of efforts to break the linear system and repurpose or upcycle waste plastic to value added products. One of the most recent examples is the chemical conversion of single-use PE to lubricants

by a catalytic upcycling process using platinum nanoparticles supported on perovskites [8,12]. Producing lubricants from PE is of particular importance because they can be successfully recycled with infinite turns, creating a circular carbon economy [13,14]. The study also demonstrates that lubricants perform as well as their commercial counterparts and the conversion of PE to lubricants is economically feasible.

Another viable option that can truly close the gap in circular economy of polyolefins is the chemical conversion of waste to feedstock that are used to produce virgin polyolefins. Polyolefin precursors are produced from steam cracking of naphtha [15]. Naphtha is a hydrocarbon fraction, usually constituents 15–30 weight % of crude oil and has a boiling point range between 30°C and 200°C. It contains hydrocarbon molecules with 5–12 carbon atoms, mostly including saturated hydrocarbons such as paraffins and naphthenes with minor compounds including olefins and aromatics [16]. There are two types of naphtha blends produced from the distillation of crude oil in the refineries: (i) Heavy naphtha which consists of mainly alkanes and cycloalkanes with a boiling point of 70 to 200°C and is used to produce aromatics [17], and (ii) Light naphtha (also known as low-boiling naphtha) which consists of mostly pentane and hexane derivatives and is fed to the steam cracker unit to produce polyolefin precursors [18].

If the objective is to obtain virgin polyolefins from the waste to achieve circularity, waste polyolefins should be converted to a compound that resembles of naphtha and fed to the cracker unit. This is only possible by chemical recycling, a process that breaks down longer polymeric chains into smaller units which can be recycled into a range of useful materials. Various chemical recycling methods, such as pyrolysis, gasification, and hydrothermal processing can be used to convert plastic wastes into gases, fuels, and other compounds. Yet, pyrolysis is a more viable choice if the intended product is a liquid that can be fed to the steam crackers. [19,20]

Pyrolysis is the thermal degradation of hydrocarbon-based feedstock materials by heating in an oxygen-free environment at high temperatures (300 – 700 °C). Because of heating, large chain polymers decompose into smaller hydrocarbons. The pressure is typically atmospheric although it can also be performed under vacuum. The pyrolysis performance also depends on the properties of the feedstock such as molecular structure including chain irregularities, branching, initiators, chain lengths and crystallinity, etc. When a pyrolysis-like technology is applied, carbon number of the cleaved polymers decrease, eventually reaching a point that it exists as liquid in pseudo-equilibrium with its vapor in the pyrolysis reactor. The properties of the liquid obtained by pyrolysis can be very similar to conventional fuels (in terms of energy content, octane and cetane number, and other physical properties such

as density, viscosity, flash-point etc.). What in turn determines the properties of these liquids is the chemical composition of the liquid (including aromatic content, distillation range, paraffinic content, etc) [21–25].

Herein, the suitability of pyrolysis as a stand-alone chemical recycling technique for producing the precursors for virgin polyolefins is examined. Liquid products obtained from noncatalytic pyrolysis of polypropylene at two different temperatures were analyzed to obtain physical properties and chemical composition by several chromatography techniques, in comparison to two different naphtha mixtures, namely, light naphtha and heavy naphtha. Liquid samples were then distilled under vacuum and fractionated to examine if a portion of pyrolysis oil can be used as naphtha. Results demonstrate that, although there is a very small fraction of pyrolysis oil consisting of saturated alkanes and naphthenes, pyrolysis oil obtained from PP exhibits distinct compositional differences than naphtha and cannot be used as a substitute for it.

Materials and Methods

Pyrolysis oils were produced by thermal non-catalytic degradation pyrolysis of polypropylene in two different temperature ranges. The polypropylene used in this study are bigbag sacks. 'Py oil-1' represents the liquid product from the reactor at 270-300°C, and 'Py oil-2' represents the liquid product taken from the reactor at 370-400°C. The values of light and heavy naphtha samples were taken as the analysis results of the naphtha products used as inputs in the Ethylene and Aromatics plants of PETKİM, respectively.

Density measurements were performed according to American Society for Testing and Materials (ASTM) D4052 - Standard Test Method for Density, Relative Density, and the American Petroleum Institute (API) Gravity of Liquids by Digital Density Meter. Vacuum distillation results were determined with ASTM D1160 - Standard Test Method for Distillation of Petroleum Products at Reduced Pressure. Total sulfur amount of the samples was measured with ASTM D453 - Standard Test Method for Determination of Total Sulfur in Light Hydrocarbons, Spark Ignition Engine Fuel, Diesel Engine Fuel, and Engine Oil by Ultraviolet Fluorescence.

To understand the boiling point distribution and the amount of the naphtha fraction of the pyrolysis oils, simulated distillation (SIMDIST) analysis ASTM D7169 were performed and boiling points were determined according to ASTM D86 - Standard Test Method for Distillation of Petroleum Products and Liquid Fuels at Atmospheric Pressure. SIMDIST is a Gas Chromatography (GC) method applied to characterize and separate petroleum fractions and products based on their boiling point. The principle of SIMDIST is based on a combination of

traps for hydrocarbons with a similar functional group, separating columns and a hydrogenator for olefins. Atmospheric distillation is associated with the volumetric composition, energy content and boiling range distribution of fuels and petroleum products.

Alkane profile of the samples were also obtained to examine carbon number distribution. Bromine number values of the samples were calculated by applying the method from ASTM D1159-07 - Standard Test Method for Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration and compared with light and heavy naphtha results.

To examine the chemical compositions, PIONA (acronym for n-paraffins, iso-paraffins, olefins, naphthenes and aromatics) analysis was performed on pyrolysis oils fractionated up to 210°C. This temperature was chosen to prevent column clogging in PIONA analysis. PIONA is a unique method used in the Refinery and Petrochemical industries. It is a multi-dimensional chromatography technology containing seven separation columns. For each mode, all settings are preprogrammed to enable automatic column switching and temperature control. The analyzer system is specified according to the following standard methods, Euopean Norm International Organization for Standardization (EN ISO) 22854 and ASTM D6839.

Results and Discussion

Two different pyrolysis oils, shown in Figure 1, were studied in comparison with light and heavy naphtha to examine suitability of using pyrolysis oil for producing plastics to close the gap in circularity of plastics. The first observed difference is the color of the samples. Py oil-1 is yellow, Py oil-2 is orange, light and heavy naphtha are colorless.

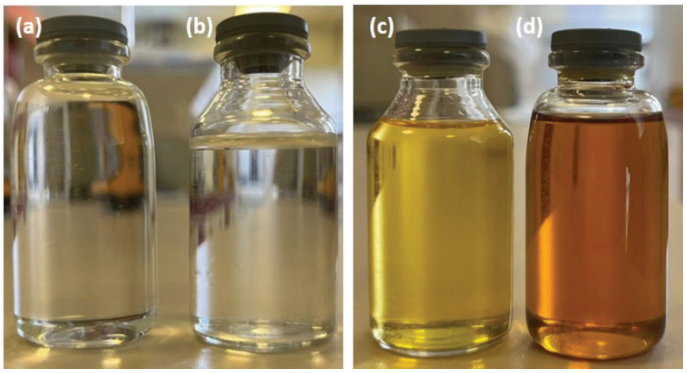

Figure 1. (a) Light naphtha, (b) Heavy naphtha, (c) Py oil-1; the oil collected between 270-300°C, (d) Py oil-2; the oil collected between 370-400°C

All four liquids, two different pyrolysis oils and naphtha mixtures have similar density values (Table 1) at 15 °C, as determined by ASTM D4052 method. Density of pyrolysis oils were closer to the typical value known for heavy naphtha. In addition to density, total sulfur amount is determined by ASTM D5453. The total sulfur content was determined to be higher in the Py oil-1 sample collected at lower temperature compared to the naphtha samples, and within the desired specification values in the Py oil-2 sample collected at higher temperature. Yet, it should be noted that the purity of pyrolysis oil depends on the purity of the waste plastics. If waste plastic is contaminated with sulfur-containing compounds, pyrolysis mixtures will contain higher amount of sulfur.

The boiling point fractionation was determined with vacuum distillation by ASTM D1160 method and tabulated in Table 1. Vacuum distillation results of light naphtha and heavy naphtha are similar, albeit slightly lower initial boiling point (IBP) of light naphtha. Vacuum distillation results of pyrolysis oils, however, show major differences. IBP of pyrolysis oils are slightly higher than naphtha mixtures. The required temperature for 90% fractionation of both oils is around 380-385 °C, significantly higher than that of naphtha mixtures.

Table 1. Density, vacuum distillation and total sulfur results

Test	Unit	Light Naphtha	Heavy Naphtha	Py oil-1	Py oil-2	Method
Density, 15°C	g/cm^3	Min. 0.67, max. 0.72	≥ 0,73	0.7917	0.7840	ASTM D4052
Total Sulfur	wt. %	Max. 0.060	Max. 0.10	0.3	0.006	ASTM D5453
IBP*	°C	≥ 33	≥ 50	60	55	
5 % Fraction	°C	-	-	97	97	
10 % Fraction	°C	-	Min. 85	115	115	
20 % Fraction	°C	-	-	145	140	
30 % Fraction	°C	-	Min. 105	155	165	
40 % Fraction	°C	-	-	195	210	ASTM D1160
50 % Fraction	°C	Min. 115	Min. 120	232	242	
60 % Fraction	°C	-	-	272	280	
70 % Fraction	°C	-	Min. 135	325	338	
80 % Fraction	°C	-	-	365	372	
90 % Fraction	°C	Min. 170	Min. 170	380	385	

* Initial boiling point.

Fractionation of pyrolysis oils and naphtha liquids was also performed with Simulated Distillation (SIMDIST) to determine the volumetric fraction of

pyrolysis oil that is similar to naphtha based on where the final boiling point (FBP) of light and heavy naphtha are located in distillation results of pyrolysis oils. This information is then used for determining the volumetric percentage. SIMDIST results in Table 2 show that light naphtha-like composition is 10-15% in Py oil-1 and 5-10% in Py-oil 2, whereas heavy naphtha-like composition is 30-40% in both Py oil-1 and Py oil-2.

Table 2. SIMDIST analysis results

Mass (%)	Boiling Point (°C)			
	Light Naphta	Heavy Naphta	Py oil-1	Py oil-2
5	36.4	96.0	61.7	66.4
10	40.8	101.5	110.1	123.5
15	-	107.6	132.5	133.2
20	46.0	113.1	133.9	134.3
30	50.7	117.9	136.0	158.9
40	55.6	125.0	190.2	227.6
50	60.6	128.7	234.5	246.9
60	66.6	134.7	275.2	303.4
70	72.2	140.5	320.0	339.7
80	79.1	146.5	374.3	382.4
85	-	150.3	407.2	405.4
90	89.9	151.8	436.2	430.4
95	100.2	157.9	474.0	464.0
96	-	159.1	485.2	472.2
97	-	160.7	495.8	483.6
98	-	163.2	508.9	497.2
99	-	167.1	527.5	517.7
FBP*	117.9	171.4	541.6	537.9

* Final boiling point.

In Figure 2, alkane profile of the samples based on the carbon numbers were illustrated. C8 fraction had the highest value in pyrolysis oil samples. These values were determined as 28.7 % for Py oil-1 and 25.6 % for Py oil-2. The carbon number distribution of pyrolysis oils varied in the range of C5-C44.

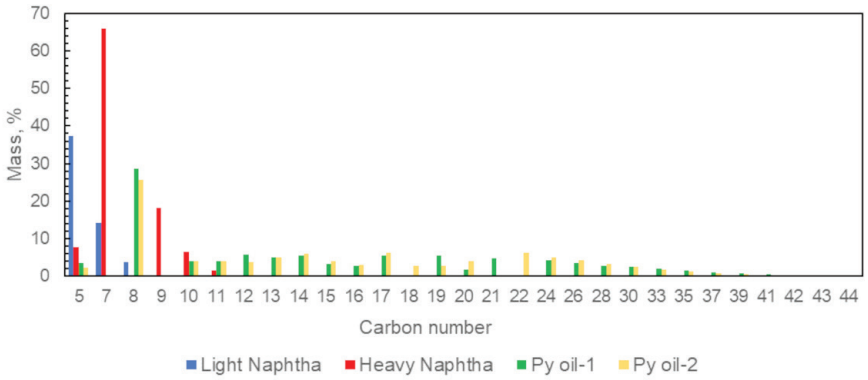

Figure 2. Alkane profile of the samples

The distillation studies and alkane distribution analysis point out that naphtha mixtures have significantly lower number of carbons than pyrolysis oils. This is a major difference between naphtha and pyrolysis oil that may limit using the latter as a substitute for the former. Existing pyrolysis technologies are, however, equipped with a condenser that helps with decreasing the carbon number to the desired range. For instance, the patent disclosed by BlueAlp Innovations B.V. technology for chemical recycling of plastics describes a use of partial condenser, which controls the composition of the pyrolyzed gas by the condenser temperature. In addition, pyrolysis oil can be distillated to obtain desired fraction for downstream operations [26].

The carbon range of pyrolysis oil could be brought to the similar range with naphtha. What remains as a huge challenge is the compositional differences. To examine the compositional differences, bromine number of each liquid was measured and given in Table 3. The number of grams of bromine that will react with 100 g of the specimen under the conditions of the test is defined as bromine number. The bromine number quantifies and indicates the aliphatic unsaturated fraction of the petroleum products. By using this method, estimation of the percentage of olefins in petroleum distillates boiling up to approximately 315°C can be obtained, albeit lower precision above the bromine number of 185 [27]. Py oil-2 has higher bromine number (304) than Py oil-1 (85), and hence, higher olefinic content. Mangest et al. reported that straight chain olefins, branched chain olefins, cyclic olefins, and diolefins have bromine number between 63-235, 58-235, 134-237, and 185-352, respectively [28]. Based on these, it can be stated that Py oil-1 is likely to be composed of straight and branched chain olefins, while the majority of Py oil-2 contains diolefins. Note that bromine numbers for light and heavy naphtha are significantly smaller than pyrolysis oils, indicating the major compositional differences in terms of olefinic content.

Table 3. Bromine number of the pyrolysis oils and naphtha samples

Sample	Bromine Number (g/100 g)
Light Naphtha	1.2
Heavy Naphtha	0.3
Py oil-1	85
Py oil-2	304
Py oil-1-F210	88
Py oil-2-F210	216

Py oil-1 and Py oil-2 were distilled up to 210°C to obtain naphtha-like fractionations and their exact compositions using PIONA. Fractionated pyrolysis oil samples are more suitable samples for PIONA analysis than their unfractionated counterparts because PIONA analysis is limited to hydrocarbons that have boiling point lower than 200°C and carbon number around C_{11} [29]. The distillates were labelled as Py oil-1-F210 and Py oil-2-F210. Bromine numbers of the fraction of pyrolysis oils up to 210°C were given in Table 3. The distillation did not change the bromine number of Py oil-1, whereas decreased that Py oil-2 from 304 to 216. It should be noted that these numbers are still higher than bromine numbers of naphtha samples.

PIONA analysis results were given in Table 4. Saturated components (naphthenes, and paraffins) of light naphtha and heavy naphtha are approximately 98% and 91%, respectively, with remaining minor components being olefins and aromatics. The ratio of the saturated components in Py oil-1-F210 and Py-oil-2-F210 are approximately 36% and 35%, respectively, significantly lower than that of naphtha blends. The majority of components in fractionated pyrolysis oils were found to be cyclic olefins (~44%). High concentration of olefinic substances is an attribute of pyrolysis oil obtained from PP, as reported in an earlier study by Kusenberg et al. [30]. High fraction of i-paraffins in pyrolysis oils is attributed to using PP as a feedstock for pyrolysis. If PE without significant branching was used as a feedstock, higher fraction of n-paraffins would be anticipated since PE is likely to decompose into linear hydrocarbons. Aromatics in pyrolysis oils formed during pyrolysis and their formation cannot be linked other polymer resins such as polystyrene in the feedstock since virgin PP is used for pyrolysis to produce oils [24,31–35].

Table 4. PIONA analysis summary

Component	Light Naphtha, % (w/w)	Heavy Naphtha, % (w/w)	Py oil-1-F210, % (w/w)	Py oil-2-F210, % (w/w)
Naphthenes	15.7	36.3	9.3	10.1
i-Paraffins	41.7	28.8	19.1	19.1
n-Paraffins	40.3	25.9	7.8	5.5
Cyclic Olefins	-	-	45.3	43.4
Olefins	0.1	0.3	15.0	16.8
Aromatics	2.2	8.8	3.5	5.2

When pyrolysis is applied to waste PP, it produced a complex mixture of hydrocarbons with low selectivity to desired products (saturated hydrocarbons) and high selectivity to highly-reactive olefins that may form deposits on the cracker walls and aromatics that are precursors for coke formation [35]. It is anticipated that operational issues would arise if these fractionated pyrolysis oils were to fed to the steam crackers directly without any upgrading. Thus, as a stand-alone technology, pyrolysis oil cannot replace and cannot be blended with naphtha, and not a viable option for closing the circularity of waste plastics.

Conclusions

This study focuses on comparing two different pyrolysis oils to light and heavy naphtha, and critically evaluates the suitability of using pyrolysis oil as a feedstock for producing polyolefins to close the gap in the circularity. Liquid samples obtained from noncatalytic pyrolysis of PP were analyzed by several standardized tests. Analyses were also performed for two different naphtha samples for comparison purposes. Results show that although some of the physical properties such as density, initial boiling point, and sulfur content of the pyrolysis oils and naphtha are similar, pyrolysis oils examined in this study exhibit two major differences than the naphtha samples: (i) Pyrolysis oils have a wider carbon distribution than naphtha. This difference however can be alleviated by integrating a condenser to the reactor or distilling the pyrolysis oil to obtain the desired carbon range. (ii) The majority of the pyrolysis oils is cyclic olefins, whereas the majority of naphtha samples are paraffinic hydrocarbons. If the studied pyrolysis oils are used as feedstock for steam crackers, excessive carbon formation may occur inside naphtha cracker and operational issues may arise because of reactive unsaturated hydrocarbons present in pyrolysis oil.

The results demonstrate that the compositional differences prevent pyrolysis oil to be used as a substitute for naphtha and a feedstock for steam crackers. Even after distillation, less than 10% of pyrolysis oil of our study exhibited naphtha-like properties. This, however, does not mean that all pyrolysis oils fall under this conclusion. Pyrolysis oil properties are heavily dependent on system parameters, operating conditions, and catalyst attributes [24,36]. It is also possible to combine pyrolysis with other chemical technologies to upgrade the properties of pyrolysis oil. Therefore, the findings of our study should not be generalized as pyrolysis oil could never be used as a feedstock for steam crackers to close the gap in the circular economy. But if pyrolysis oil will be used as a feedstock for existing steam crackers and refinery infrastructure, it should exhibit naphtha-like properties. For making such an assessment, this study can serve as a benchmark for evaluating naphtha-like feedstock properties for steam crackers.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. US Environmental Protection Agency Advancing Sustainable Materials Management: 2014 Fact Sheet. *United States Environmental Protection Agency, Office of Land and Emergency Management, Washington, DC 20460* 2016, 22.
2. Hopewell, J.; Dvorak, R.; Kosior, E. Plastics Recycling: Challenges and Opportunities. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2009, *364*, 2115–2126, doi:10.1098/rstb.2008.0311.
3. Rahimi, A.; García, J.M. Chemical Recycling of Waste Plastics for New Materials Production. *Nat Rev Chem* 2017, *1*, 0046, doi:10.1038/s41570-017-0046.
4. Heller, M.C.; Mazor, M.H.; Keoleian, G.A. Plastics in the US: Toward a Material Flow Characterization of Production, Markets and End of Life. *Environmental Research Letters* 2020, *15*, 94034, doi:10.1088/1748-9326/ab9e1e.
5. Kunwar, B.; Cheng, H.N.; Chandrashekaran, S.R.; Sharma, B.K. Plastics to Fuel: A Review. *Renewable and Sustainable Energy Reviews* 2016, *54*, 421–428, doi:https://doi.org/10.1016/j.rser.2015.10.015.
6. The Environmental Impact of Plastic Straws – Facts, Statistics, and Infographic Available online: https://get-green-now.com/environmental-impact-plastic-straws (accessed on 11 November 2022).
7. Geyer, R.; Jambeck, J.R.; Law, K.L. Production, Use, and Fate of All Plastics Ever Made. *Sci Adv* 2017, *3*, e1700782, doi:10.1126/sciadv.1700782.
8. Celik, G.; Kennedy, R.M.; Hackler, R.A.; Ferrandon, M.; Tennakoon, A.; Patnaik, S.; Lapointe, A.M.; Ammal, S.C.; Heyden, A.; Perras, F.A.; et al. Upcycling Single-Use Polyethylene into High-Quality Liquid Products. *ACS Cent Sci* 2019, *5*, 1795–1803, doi:10.1021/acscentsci.9b00722.
9. Ragaert, K.; Delva, L.; van Geem, K. Mechanical and Chemical Recycling of Solid Plastic Waste. *Waste Management* 2017, *69*, 24–58, doi:https://doi.org/10.1016/j.wasman.2017.07.044.
10. Miandad, R.; Barakat, M.A.; Aburiazaiza, A.S.; Rehan, M.; Nizami, A.S. Catalytic Pyrolysis of Plastic Waste: A Review. *Process Safety and Environmental Protection* 2016, *102*, 822–838, doi:https://doi.org/10.1016/j.psep.2016.06.022.

11. Hong, M.; Chen, E.Y.-X. Chemically Recyclable Polymers: A Circular Economy Approach to Sustainability. *Green Chem.* 2017, *19*, 3692–3706, doi:10.1039/C7GC01496A.
12. Delferro, M.; Ferrandon, M.S.; Kennedy, R.M.; Celik, G.; Hackler, R.A.; Poeppelmeier, K.R.; Sadow, A.D. Catalytic Upcycling of Polymers 2021.
13. Hackler, R.A.; Vyavhare, K.; Kennedy, R.M.; Celik, G.; Kanbur, U.; Griffin, P.J.; Sadow, A.D.; Zang, G.; Elgowainy, A.; Sun, P.; et al. Synthetic Lubricants Derived from Plastic Waste and Their Tribological Performance. *ChemSusChem* 2021, *14*, 4181–4189, doi:10.1002/cssc.202100912.
14. Celik, G.; Vyavhare, K.; Kennedy, R.M.; Hackler, R.A.; Erdemir, A.; Delferro, M. Catalytic Upcycling of Polyolefins into Lubricants 2022.
15. Karaba, A.; Dvořáková, V.; Patera, J.; Zámostný, P. Improving the Steam-Cracking Efficiency of Naphtha Feedstocks by Mixed/Separate Processing. *J Anal Appl Pyrolysis* 2020, *146*, 104768, doi:10.1016/j.jaap.2019.104768.
16. Rahimpour, M.R.; Jafari, M.; Iranshahi, D. Progress in Catalytic Naphtha Reforming Process: A Review. *Appl Energy* 2013, *109*, 79–93, doi:10.1016/j.apenergy.2013.03.080.
17. Ross, J.R.H. Catalysis in the Production of Energy Carriers From Oil. In *Contemporary catalysis: fundamentals and current applications*; Marinakis, K., Ed.; Elsevier, 2019; pp. 233–249.
18. Speight, J.G. 9 - Gas Condensate. In *Natural gas: a basic handbook*; Hammon, K., Ed.; Gulf Professional Publishing, 2019; pp. 325–358.
19. Lange, J.-P. Managing Plastic Waste—Sorting, Recycling, Disposal, and Product Redesign. *ACS Sustain Chem Eng* 2021, *9*, 15722–15738, doi:10.1021/acssuschemeng.1c05013.
20. Das, P.; Tiwari, P. The Effect of Slow Pyrolysis on the Conversion of Packaging Waste Plastics (PE and PP) into Fuel. *Waste Management* 2018, *79*, 615–624, doi:https://doi.org/10.1016/j.wasman.2018.08.021.
21. Nanda, S.; Berruti, F. Thermochemical Conversion of Plastic Waste to Fuels: A Review. *Environ Chem Lett* 2021, *19*, 123–148, doi:10.1007/s10311-020-01094-7.
22. Dogu, O.; Pelucchi, M.; Van de Vijver, R.; Van Steenberge, P.H.M.; D'hooge, D.R.; Cuoci, A.; Mehl, M.; Frassoldati, A.; Faravelli, T.; Van Geem, K.M. The Chemistry of Chemical Recycling of Solid Plastic Waste via Pyrolysis and Gasification: State-of-the-Art, Challenges, and Future Directions. *Prog Energy Combust Sci* 2021, *84*, 100901, doi:10.1016/j.pecs.2020.100901.
23. Zhang, F.; Zhao, Y.; Wang, D.; Yan, M.; Zhang, J.; Zhang, P.; Ding, T.; Chen, L.; Chen, C. Current Technologies for Plastic Waste Treatment: A Review. *J Clean Prod* 2021, *282*, 124523, doi:10.1016/j.jclepro.2020.124523.
24. Al-Salem, S.M.; Antelava, A.; Constantinou, A.; Manos, G.; Dutta, A. A Review on Thermal and Catalytic Pyrolysis of Plastic Solid Waste (PSW). *J Environ Manage* 2017, *197*, 177–198, doi:https://doi.org/10.1016/j.jenvman.2017.03.084.
25. Lopez, G.; Artetxe, M.; Amutio, M.; Bilbao, J.; Olazar, M. Thermochemical Routes for the Valorization of Waste Polyolefinic Plastics to Produce Fuels and Chemicals. A Review. *Renewable and Sustainable Energy Reviews* 2017, *73*, 346–368, doi:https://doi.org/10.1016/j.rser.2017.01.142.
26. Van Der Ree, T.; Weser, G. Method and System for Transferring Plastic Waste into a Fuel Having Properties of Diesel/Heating Oil 2018.
27. ASTM Standard D7169-05, "'Standard Test Method for Bromine Numbers of Petroleum Distillates and Commercial Aliphatic Olefins by Electrometric Titration.'" 2008, 213–221, doi:10.1520/mnl10860m.
28. Mangesh, V.L.; Padmanabhan, S.; Ganesan, S.; Prabhudevrahul, D.; Kumar Reddy, T.D. Prospects of Pyrolysis Oil from Plastic Waste as Fuel for Diesel Engines: A Review. *IOP Conf Ser Mater Sci Eng* 2017, *197*, doi:10.1088/1757-899X/197/1/012027.
29. Lissitsyna, K.; Huertas, S.; Quintero, L.C.; Polo, L.M. PIONA Analysis of Kerosene by Comprehensive Two-Dimensional Gas Chromatography Coupled to Time of Flight Mass Spectrometry. *Fuel* 2014, *116*, 716–722, doi:10.1016/j.fuel.2013.07.077.

30. Kusenberg, M.; Zayoud, A.; Roosen, M.; Thi, H.D.; Abbas-Abadi, M.S.; Eschenbacher, A.; Kresovic, U.; de Meester, S.; van Geem, K.M. A Comprehensive Experimental Investigation of Plastic Waste Pyrolysis Oil Quality and Its Dependence on the Plastic Waste Composition. *Fuel Processing Technology* 2022, *227*, 107090, doi:https://doi.org/10.1016/j.fuproc.2021.107090.
31. Sojak, L.; Kubineca, R.; Jurdakovaa, H.; Hajekovab, E.; Bajus, M. GC-MS of Polyethylene and Polypropylene Thermal Cracking Products. *Pet Coal* 2006, *48*, 1 – 14.
32. Kopinke, F.D.; Zimmermann, G.; Reyniers, G.C.; Froment, G.F. Relative Rates of Coke Formation from Hydrocarbons in Steam Cracking of Naphtha. 2. Paraffins, Naphthenes, Mono-, Di-, and Cycloolefins, and Acetylenes. *Ind Eng Chem Res* 1993, *32*, 56–61, doi:10.1021/ie00013a009.
33. Kopinke, F.D.; Zimmermann, G.; Reyniers, G.C.; Froment, G.F. Relative Rates of Coke Formation from Hydrocarbons in Steam Cracking of Naphtha. 3. Aromatic Hydrocarbons. *Ind Eng Chem Res* 1993, *32*, 2620–2625, doi:10.1021/ie00023a027.
34. Soják, L.; Kubinec, R.; Jurdáková, H.; Hájeková, E.; Bajus, M. High Resolution Gas Chromatographic–Mass Spectrometric Analysis of Polyethylene and Polypropylene Thermal Cracking Products. *J Anal Appl Pyrolysis* 2007, *78*, 387–399, doi:https://doi.org/10.1016/j.jaap.2006.09.012.
35. Kusenberg, M.; Faussone, G.C.; Thi, H.D.; Roosen, M.; Grilc, M.; Eschenbacher, A.; de Meester, S.; van Geem, K.M. Maximizing Olefin Production via Steam Cracking of Distilled Pyrolysis Oils from Difficult-to-Recycle Municipal Plastic Waste and Marine Litter. *Science of The Total Environment* 2022, *838*, 156092, doi:https://doi.org/10.1016/j.scitotenv.2022.156092.
36. Dogu, O.; Pelucchi, M.; van de Vijver, R.; van Steenberge, P.H.M.; D'hooge, D.R.; Cuoci, A.; Mehl, M.; Frassoldati, A.; Faravelli, T.; van Geem, K.M. The Chemistry of Chemical Recycling of Solid Plastic Waste via Pyrolysis and Gasification: State-of-the-Art, Challenges, and Future Directions. *Prog Energy Combust Sci* 2021, *84*, 100901, doi:https://doi.org/10.1016/j.pecs.2020.100901.

High expression of ring-hydroxylating dioxygenase genes ensure efficient degradation of p-toluate, phthalate, and terephthalate by Comamonas testosteroni strain 3a2

Didem Aksu[1], Mamadou Malick Diallo[2], Umut Şahar[3], Tayyibe Alpay Uyanıker[4], Guven Ozdemir[3]

[1]*Application and Research Center for Testing and Analysis, Ege University, Izmir, Türkiye*
[2]*Environmental Science Department, Faculty of Science, Ege University, Izmir, Türkiye*
[3]*Biology Department, Faculty of Science, Ege University, Izmir, Türkiye*
[4]*SOCAR, Türkiye R&D and Innovation Co, Aliağa, Izmir, Türkiye*

Journal: Archives of Microbiology, **2021**, 203, 4101–4112

https://link.springer.com/article/10.1007/s00203-021-02395-3

DOI: https://doi.org/10.1007/s00203-021-02395-3

Abstract

Para-toluic acid, a major pollutant in industrial wastewater, is hazardous to human health. It has been demonstrated that Gram-negative bacteria are among the most effective degraders of para-toluic acid. In this study, the ability of Comamonas testosteroni strain 3a2, isolated from a petrochemical industry wastewater, to degrade para-toluic acid was investigated. The effect of different carbon (glucose and ethylene glycol) and nitrogen sources (urea, yeast extract, peptone, $NaNO_3$, NH_4NO_3) on the biodegradation of para-toluic acid by the isolate 3a2 was evaluated. Furthermore, ring hydroxylating dioxygenase genes were amplified by PCR and their expression was evaluated during the biodegradation of para-toluic acid. The results indicated that strain 3a2 was able to degrade up to 1000 mg/L of para-toluic acid after 14 h. The highest degradation yield was recorded in the presence of yeast extract as nitrogen source. However, the formation of terephthalic acid and phthalic acid was noted during para-toluic acid degradation by the isolate 3a2. Toluate 1,2-dioxygenase, terephthalate 1,2 dioxygenase, and phthalate 4,5 dioxygenase genes were detected in the genomic DNA of 3a2. The induction of ring hydroxylating dioxygenase genes was proportional to the concentration of each hydrocarbon. This study showed that the isolate 3a2 can produce terephthalate and phthalate during the para-toluic acid biodegradation, which were also degraded after 24 h.

Keywords: Ring Hydroxylating Dioxygenases;Para-Toluic acid; Toluate 1,2-dioxygenase; terephthalate 1,2 dioxygenase; phthalate 4,5 dioxygenase; biodegradation

Introduction

Para-Toluic acid (pTol), also known as 4-methylbenzoate, is widely generated during the production of purified terephthalic acid (PTA). The amount of pTol in wastewateris estimatedabout 30% of the total chemical oxygen demand (COD) produced after PTA production(Noyola et al.2000). Besides, pTol is used as a raw material for the production of anticorrosive additives, colourants, dyestuffs and paints(Maki and Takeda 2000). However,pTol is very harmful to humans due toitshazardous and toxic nature. It is recognized as hazardous waste by the United States Environmental Protection Agency (USEPA) and is categorized in the priority pollutant list(Shirota et al.2008). Taking into consideration the toxic and carcinogenic effects of pTol, it is necessary to establish environmentally friendly, cost-effective and efficient techniques to ensure the safe removal of pTol from polluted sites. Microbial degradation represents an important approach for eliminating hydrocarbons from contaminated ecosystems. In this context, bioremediation has been recognized as a recommended method for removing pTol.Previous studies showed that optimizing bacterial culture media can enhance the biodegradation process (Abd-El-Haleem et al. 2003). The presence of different carbon and nitrogen sources in culture media plays an important role during the biodegradation process (Varjani and Upasani 2016, Varjani and Upasani2017; Zhao et al.2011).

Various bacterial strains belonging to *Bacillus, Pseudomonas, Burkholderia, Acinetobacter*, *Comamonas, Mycobacterium, Rhodococcus* genera have been reported as pTol degraders(Phale et al. 2007). These microorganisms degrade aromatic compounds aerobically through key multicomponent enzymes named ring hydroxylating dioxygenases (RHDs), which are also known as Rieske non-heme iron dioxygenases. RHDs consist of twoα_n/β_nsubunits and contain a Rieske-type Fe_2S_2 center.The active-site mononuclear iron is involved in the initial oxidation of the aromatic compounds(Vermaet al. 2019). Toluate 1,2-dioxygenase (TADO) and benzoate dioxygenase (BADO) are toluate dioxygenases which catalyze the 1,2-dihydroxylation of different kind of benzoates. TADO transforms wider range of substituted benzoates when compared to BADO. The α/β subunits of TADO and BADO are encoded by *xylXY* and *benAB* genes respectively(Ge *et al.* 2002; Ge and Eltis 2003). However, Junkeret al.(1997) described another pathway that mediated the degradation of pTol by two *Comamonas* strains (PSB-4 and T-2). These strains degrade pTol to terephthalic acid (TPA)inthree-stepprocessusing a monooxygenase (TsaMB) and two dehydrogenase enzymes (TsaC and TsaD), which are encoded by *tsaMBCD* genes respectively (Junker *et al.* 1997).

In the present study, an efficient aromatic hydrocarbon-degrading *Comamonas testosteroni* strain 3a2,isolated from a petrochemical industry wastewater, was used for pTol degradation. The degradation pathways and the expression of ring hydroxylating dioxygenase genes were evaluated during the degradation process.

Materials and Methods

Chemicals and Culture media

All bacterial culture media, p-Toluic acid (pTol), Bushnell Haas (BH) medium, sugar standard D-(+) glucose, ethylene glycol (EG), and acetonitrile were purchased from Merck Company (Germany). Ultrapure water was obtained using Sartorius 610 (Hampton, NH, U.S.A).

Isolation of bacterial strain

To use and publish legally the results, appropriate authorizations have been obtained from the responsible authorities of SOCAR company The 3a2 isolate used in this study was isolated from a petrochemical industry wastewater in SOCAR Co, Izmir, Türkiye (Project number: 112D044-TUBİTAK).The isolation of bacterial strain was carried out in a flask containing 50 mL Bushnell-Hass medium (BH) (g L^{-1}: KH_2PO_4, 1.00; K_2HPO_4, 1.00; $MgSO_4$, 0.2; $CaCl_2$, 0.02; NH_4NO_3, 1.00; $FeCl_3$, 0.05; yeast extract, 0.05) inoculated with 2.5 g sludge and supplemented with 50 mgL^{-1}pTol as sole carbon source.The flask was incubated at 30°C at 150 g for 7 days (Zhang et al. 2013). Then, 2.5 mL of the enriched culture were transferred to a new flaskcontaining fresh BH mediumsupplemented with 50 mgL^{-1} pTol and incubated under the same conditions. To obtain the enriched pTol degrading bacteria,this process was repeated four times. At the end of incubation, 1mL from the flask was serially diluted using sterile 1x phosphate-buffered saline (1x PBS, pH 7.4).TwentyµL from the 10^3dilution were transferred to BH agar supplemented with 50 mgL^{-1} pTol. The agar plates were incubated at 30°C for 3 days. The colonies were further cultured on Nutrient Agar (NA) (Merck, Germany) plates and incubated at 30°C for 2 days. Then, single colonies were transferred to sterile NA plates. The inoculated plates were incubated under the same conditions as mentioned above. The obtained pure colonies were then screened for their pTol degradation ability.

Screening of pTol biodegradation

Each isolate was individually inoculated into 5 mL nutrient broth (g L^{-1}: beef extract, 1.0; peptone, 5.0; yeast extract, 2.0; NaCl, 5.0) and incubatedovernight in a shaking incubator (150 g) at 30°C. One mL from each sample was centrifuged at 5000 g for 5 minutes and the supernatants were discarded. The pellets were washed twice with sterile 1x PBS, and the concentration of the bacterialcells in each samplewas fixed to 1.5 x 10^8CFU/mL using the DEN-1B McFarland Densitometer device (BioSan Densitometer GmbH, Germany). Thenflasks containing 50 mL BH medium supplemented with 50 mgL^{-1} pTol were inoculated with 2.5 mL from each standardized culture.The flasks were incubated in a rotary shaker (150 g) at 30°C for 24h. A flask that contained

un-inoculated sterile medium was employed as control. The pTol degradation yields were assessed every two hours for 24h.The isolate that demonstrated the highest pTol degradation was selected for further experiments. The isolate was maintained in Nutrient Broth containing 16 % glycerol and stored at -80°C.

All assays were conducted in triplicate. Unless otherwise indicated, all the experiments in the study were performed as mentioned above.

Molecular characterization of 3a2 isolate

The genomic DNA (gDNA)of the strain 3a2 was extracted using the Presto Mini gDNA Bacteria Kit (Geneaid, Taiwan) according to the manual's instructions. The concentration and purity of the gDNA were assessed using Nanodrop 2000c spectrophotometer (Thermo Scientific, USA). The universal primers listed in table 1 were employed to amplify the 16S rDNA by PCRusing the Q5 High-Fidelity DNA Polymerase (BioLabs, USA).The PCR amplification was carried out using a TechnePlus Thermal cyclerand the amplification stages were as follow: Initial denaturation stage: 1 cycle of 95°C for 120 s;amplification stage (30 cycles): 95°C for 20 s, 56°C for 40 s, and 72°C for 90 s; final extension stage: 72°C for 5 min.The PCR productswere visualized on 1% agarose gel using Bioimaging Systems (UVP Biospectrum, UK). The 1481 bp PCR fragments were purified using Expin Combo GP kit (GeneAll, Portugal) andsequenced by MedSanTek Lab Company. The obtained nucleic acid sequences were edited using DNA Baser Assembler software and analyzedusingthe BLAST/NCBI platform.The 16S rDNA sequenceswere aligned with reference sequences using MEGA 7 software.

Table 1. List of primers used for polymerase chain reaction

Primer target	Primer names	Nucleotide sequences	Product size (bp)
16S rRNA	11F [a] (Zhang et al. 2015) 1492R	GTTTGATCCTGGCTCAG GGYTACCTTGTTACGACTT	1480
	F341[b] (De Mandal et al.2015) R518	CCTACGGGAGGCAGCAG ATTACCGCGGCTGCTGG	184
TADO	ToLDF[a] ToLDR	AACGACTATTACACCACGCAGA CATGCCGCTGACATTGAAGAAG	900
	qpTolF[b] qpTolR	GAGAGAAGCGGGCGATGATA CCTCACTGGCTAAGCGATCT	165
TPADO	TPaDF[a] TPaDR	TAYCACGCCTGGAGCTACAA GTTCTGRATCTGCTGCAGCAC	500
	qTpaDF[b] qTpaDR	AGCTGAATCGCCTCTCACAAA CGAACTCCTCAAAGCCCTCTA	183
PDO	ophA2F[a] ophA2R	AAGAAAACGAATTGCTGTGC ACGTTGTACAGGTTGTTCGG	750
	qophA2F[b] ophA2R	AGCCTGCACTCGTCGGACTT ACGTTGTACAGGTTGTTCGG	242
BADO	BnDOF[a] BnDOR	TACCTGGCMCAYGARAGCCA AAGAAATCYTCRTACTGGCG	900
ANTDO	AntDOF[a] AntDOR	TGGATCTACGCCTGCCACGAAA GTGGCCACGTAGTTGTAGTG	541
TsaM	TsaMF TsaMR	CTGTACCACGGCCTGAAGTT TAGATCTTCTCGGTGAGGGC	665
TsaD	TsaDF TsaDR	AACTTTCCCCTGGTGCTCTC CTCTTCGCTCATCACTGGC	678

[a]Used for PCR

[b] Used for qPCR

pTol degradation assays

The degradation efficiency of pTol by isolate 3a2 was evaluated in the presence of different pTol concentrations as well in different carbon and nitrogen sources. The pTol degradation yields were assessed every two hours for 24 hours. The capability of strain 3a2 to degrade high pTol concentrations was investigated in 50 mL BH medium supplemented with 50, 100, 250, 500 and 1000 mgL^{-1} pTol. The effect of (different) nitrogen sources on the biodegradation of pTol by 3a2 isolate was evaluated in 50 mL BH medium supplemented with 0.05% of either organic (yeast extract, peptone and urea) or inorganic (sodium nitrate, ammonium nitrate) nitrogen sources.

The influenceof different carbon sources on the pTol degradation process was investigated by adding 50, 100, 500, 1000, and 2000 mgL^{-1} of glucose or EG into 50 mL BH in presence of 500 mgL^{-1}pTol.

High-Performance Liquid Chromatography (HPLC) analysis

The degradation of pTol, TPA and PAwasanalyzedusing high-performance liquid chromatography (HPLC, Agilent 1100 Series, USA) equipped with a DAD detector. HPLC analyses were carried out according toPillai *et al.*2009 with minor modifications; the mobile phase consisted of amixture of water (A) and acetonitrile/0.1% Formic acid (B) (A=70% - B= 30%) with a flow rate of 1 mL/min. AnAgilent Zorbax Eclipse PAH Column was used as astationary phase and the column temperature was 25°C. The volume of samples injected was1 µL. The detection of pTol, TPA and PAwas monitored at 240, 254 and 233 nmwith retention time equivalent to 6.2, 1.65 and 1.53 min respectively.

An HPLC equipped with a RID detector (Agilent Technologies, 1200 series, Santa Clara, CA, USA)was used to determine the capability of the isolate to use glucose and ethylene glycol (EG) as carbon sources. The separation was achieved on the Hi-Plex H column (7.7x300 mm, 8 µm) (Agilent Technologies, Santa Clara, CA, USA) under the isocratic condition with 0.0005 M H_2SO_4 as mobile phase. The flow rate was optimized and fixed at 0.600 mL/min for analysis of glucose and ethylene glycol. The column temperature was maintained at 45°C Throughout the analysis process. Detector temperature was isothermal at 45°C. The volume of injected sample was equivalent to 100 µL.

Identification of Aromatic Ring-Hydroxylating Dioxygenase Genes

The presence of seven aromatic ring-hydroxylating dioxygenase genes namely *toluate 1,2 dioxygenase (TADO), benzoate 1,2-dioxygenase (BADO), phthalate 4,5 dioxygenase (PDO), terephthalate 1,2 dioxygenase (TPADO), anthranilate 1,2-dioxygenase (ANTDO), p-toluene sulfonate monooxygenase (TsaM)* and *4-CBA dehydrogenase (TsaD)* in the gDNA of the isolate 3a2 were investigated. The specific primers that were employed to amplify the genes of interest are mentioned in Table 1. The PCR amplification was carried out in the same run for all genes as follows: initial denaturation stage 3 min at 95°C; amplification stage (35 cycles) 30 s. at 95°C, 30 s. at 52°C and 1 min at 72°C; final extension stage 5 min at 72°C. The PCR products were visualized on 1.5 % agarose gel using a multispectral imaging system (BioSpectrum, UK). The expected PCR fragments were purified using the Expin Combo GP kit (GeneAll, Portugal) and sequenced using an automated sequencer ABI Prism Genetic Analyzer (Applied Biosystems, USA). The nucleotide sequences were edited with DNA Baser Sequence Assembler software and analysed using the NCBI Blastn and Blastx search tools. Phylogenetic trees were performed with MEGA7 software using reference amino acid sequences downloaded from GenBank and UniProt databases.

Total RNA isolation

Total RNA isolation was proceeded attime zero, as well as after incubation for 24,6,8 and 10h using the Presto Mini RNA Bacteria Kit (Geneaid, Taiwan) according to manual's instructions. Before RNA isolation, aliquots from the flasks werecentrifuged at 5000 g for 5 minutes (4 °C). The cell pellets were suspended in 1x PBS and the concentration of the cells was fixed to 0.5 MacFarland using DEN-1B McFarland Densitometer device (BioSan Densitometer GmbH, Germany). Then, 1 mL aliquot from each standardized suspension wascentrifuged at 5000 g for 5 minutes (4 °C), and total RNA was isolated from the cell pellet. The concentration and purity of nucleic acid were determined using NanoDrop™ 2000c Spectrophotometer (Thermo Scientific, USA).

Reverse Transcription (RT) - qPCR analysis

The expression levels of the aromatic ring-hydroxylating dioxygenase genes were evaluated during the degradation of pTol by the selected isolate. Total RNAs were transcribed into cDNA using High Capacity cDNA Reverse Transcription Kit (AppliedBiosystems, USA) according to manual's instructions. The reverse transcription was carried out in a T100 thermal cycler (Bio-Rad, USA) as follows: 10 min at 25°C; 120 min at 37°C and 5 min at 85°C. The cDNAs were stored at -20°C. Real-time PCR experiments were performed in a Roche LightCycler® 96 Real-time PCR System (Roche Diagnostics GmbH, Germany) using LightCycler FastStart Master SYBR Green I Kit (Roche GmbH, Germany). Reaction conditions were set as recommended by the manufacturer. All genes were amplified in the same run. The real-time PCR amplification was carried out as follows: Initial denaturation stage 95°C for 10 min; Amplification stage (45 cycles): 95°C for 30 s., 51°C for 10 s. and 72°C for 10 s.; Melting stage from 65°C to 95°C at 0.2°C/s melt rates and cooling stage 40°C for 10 s. Gene expression analyses were performed using the comparative critical threshold, $2^{-\Delta\Delta CT}$ Method (relative quantification) (Livak and Schmittgen 2001). The ribosomal gene 16S rRNAwas employed as a reference gene for normalizing gene expression data.

Statistical analysis

Statistical analysis was proceeded using SPSS Statistics V.25.0 (IBM SPSS Statistics, USA). Shapiro-Wilk test was used for normality of quantitative data analysis. Repeated measures ANOVA method was used to compare the time-dependent variation of pTol biodegradation yields and growth of 3a2 in the presence of different carbon and nitrogen sources. Time- dependent pTol degradation change was compared using Tukey HSD analysis with yeast extract as a reference nitrogen source.Bonferroni correction for the number of correlations was examined.One-way ANOVA method was used for

comparisons between groups. Tukey HSD method was used mostly for binary comparisons. p-value <0.05 was considered to be significant.All experiments were performed in triplicate and the results were expressed as an average of three independent trials ± standard deviation.

Results

Identification of the strain 3a2

Phylogenetic analysis of isolate 3a2 (accession number: MT127003) indicated high 16S rDNA nucleic acid sequence similarity with *Comamonas testosteroni*strains.

Biodegradation of pTol

Strain 3a2 demonstrated high ability to use up to 1000 mgL^{-1} pTol as a carbon source (Fig. 1).

High concentrations of hydrocarbons can be toxic for microorganisms. Aromatic hydrocarbons accumulate in the lipid bilayer membrane of the cells which cause theirdisruption (Kusumawardhani et al.2018; Heipieper and Martinez 2010). The toxic effects of hydrocarbons on microorganisms have important consequences on the biodegradation process. Therefore, the addition of microorganisms capable of degrading high concentrations of toxic pollutants might accelerate their removal. In this study, the strain 3a2 was able to degrade up to 1000 mgL^{-1} pTol. The formation of PA and TPA as a result of pTol degradation was detected. The occurrence of PA and TPA startedafter 4h and their concentrations increased to reach a maximum equivalent to 93 ± 9 mgL^{-1} and 107 ± 3 mgL^{-1}respectively after 8h. While all the 500 mgL^{-1}pTol were consumed after 10h, TPA and PA were completely degraded after 12h and 24h respectively (Fig. 2).

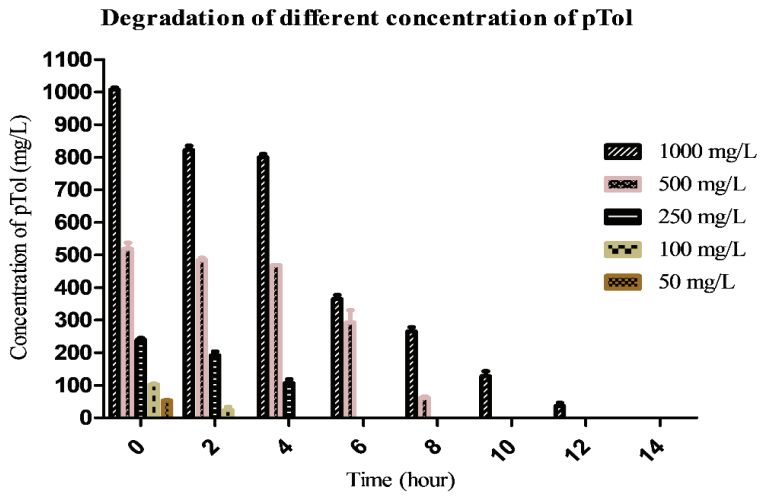

Figure 1. Degradation of different concentrations of pTol

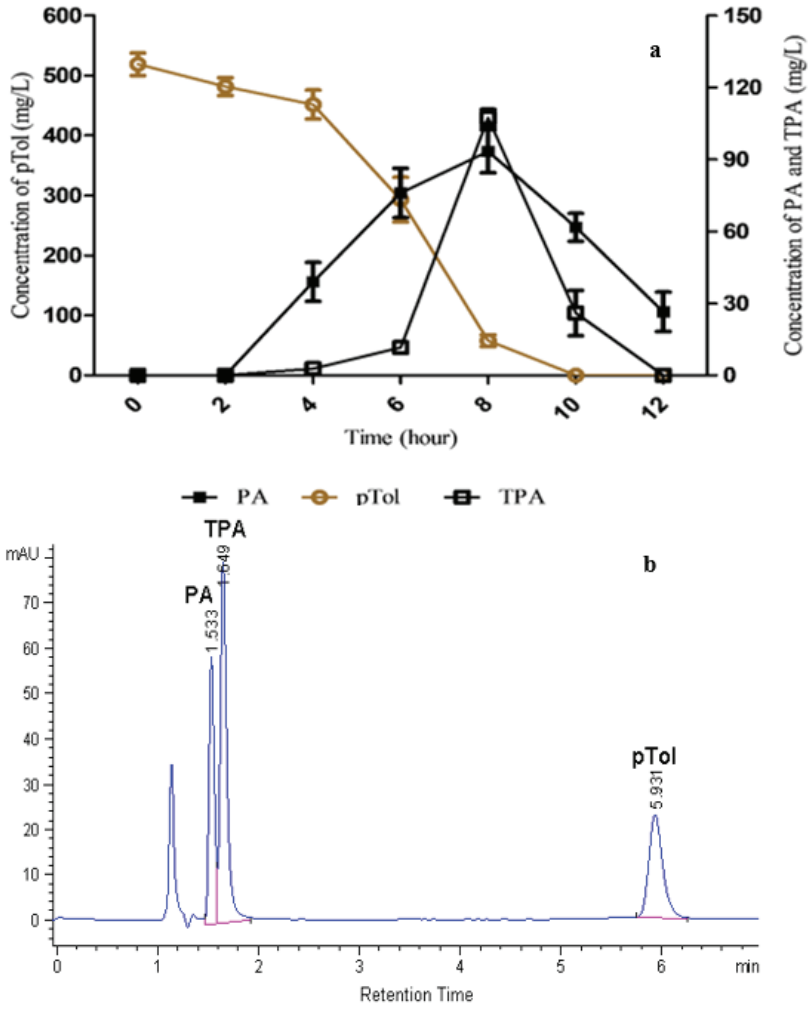

Figure 2. Formation of PA and TPA during pTol degradation (a) and HPLC image of degradation process (b)

Effect of nitrogen sources on the biodegradation of pTol

The highest degradation yield was recorded in the presence of yeast extract as 500mgL^{-1} pTol was completely removed after 10h (Fig.3a). The degradation yields in the presence of peptone and urea were 77 ± 0.47% and 40 ± 2.08%, respectively,after 12h. On the other hand, the lowest degradation yields ofpTol were obtained in the presence of inorganic nitrogen sources with yieldsequivalent to37± 3.7% and 43 ± 4.5% for sodium nitrate and ammonium nitrate respectively after 12h.The time-dependent variation of pTol concentration in the presence of different nitrogen sources was investigated with repeated measures ANOVA method. The decrease in pTol concentration with time in presence of different nitrogen sources was not similar (time-nitrogen interaction $p<0.05$). In addition, the pTol biodegradation within each nitrogen source overtime was evaluated. The variation of pTol concentration in each nitrogen source overtime were found statistically different ($p<0.001$ for each nitrogen source).

Variation of pTol concentration with time change was compared with Post-hoc test and Tukey HSD analysis.Taking into consideration that yeast extract was considered as a standard, the degradation of pTol in presence of different nitrogen sources was significantly affectedafterthe 4th h ($p <0.05$) (Supplementary Table1).The growth of *Comamonas testosteroni*strain 3a2 in the presence of different nitrogen sources was also examined. The highest growth of 3a2 was recorded in the presence of yeast extract and peptone (Fig.3b).The effect of different nitrogen sources on bacterial growth was significant. The highest difference in growth was noted after 4h ($p<0.05$).The results revealed that the adaptation of strain 3a2 to the high toxic concentrations of pTol required 4h.

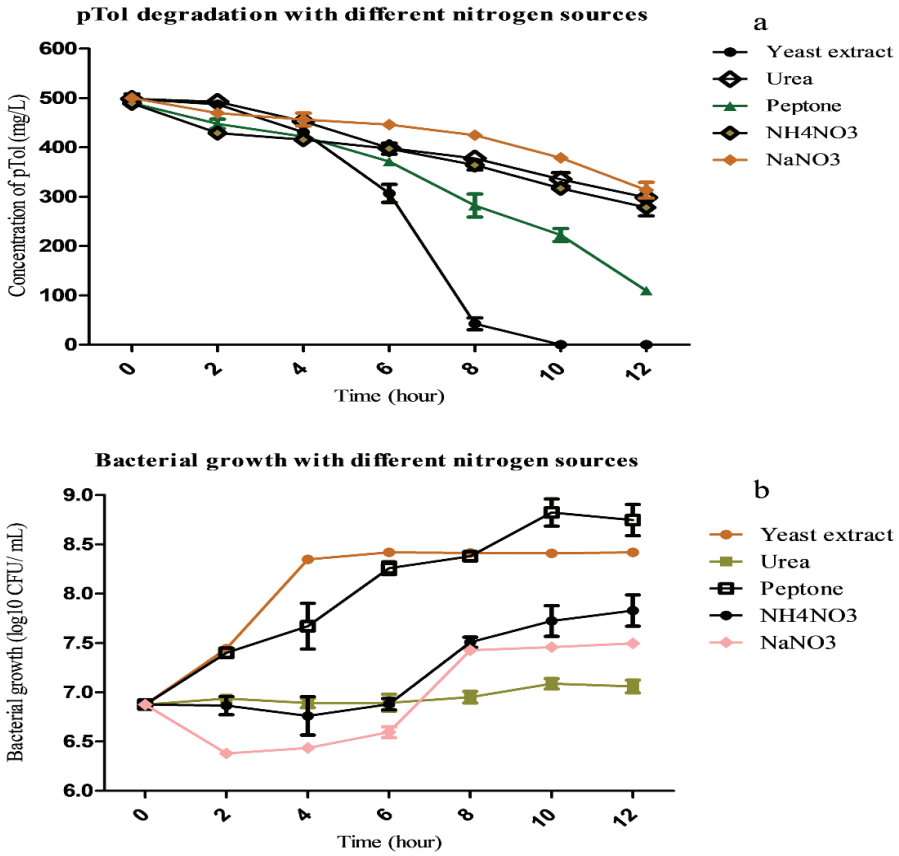

Figure 3. pTol degradation (a) and bacterial growth (b) with nitrogen sources

Biodegradation of pTol in the presence of glucose and ethylene glycol

Previous studies have investigated the impact of various carbon sources on hydrocarbon degradation by different microorganisms(Guo et al.2014; Xie et al.2013). This study investigated the effect of glucose on the degradation of pTol byisolate 3a2. As shown in Fig. 4a, glucose had no significant effecton the degradation yield.pTol was completely degraded in the presence of different concentrations of glucose (50 mgL^{-1}, 100 mgL^{-1}, 500 mgL^{-1}, 1000 mgL^{-1}, 2000 mgL^{-1}) after 12h (Supplementary Table 2). The variation of pTol concentration with time, in the presence of differentconcentrations of glucose, was investigatedusing repeated measures ANOVA method. Nosignificant difference was noted within the first 4h ($p > 0.05$). The concentration of pTol in the presence of different concentrations of glucose was compared to that in control sample (without glucose) using the Tukey HSD method. The difference in pTol concentration was significant only in the presence of 1000 mgL^{-1} and 2000mgL^{-1} glucose after 8h (Supplementary Table 3).Therefore, the presence of glucose as carbon source did not influence the pTol degradation by the isolate 3a2.

Ethylene glycol (EG) is a toxic compound that is widely used with TPA for PET production. It is extracted from natural gases and can be found in high quantity in petroleum industrial plant.In our study, the effect of different concentrations of EG on the biodegradation of pTol by strain 3a2 was investigated. According toFig. 4b,no inhibitory effect was noted in the presence ofEG at low concentrations (50 and 100 mgL^{-1}). On the other hand, it was noted that high concentrations of EG (500, 1000, 2000 mgL^{-1}) inhibited the biodegradation ability of strain 3a2. The ability of 3a2 to degrade EG as a sole carbon source was also evaluated using HPLC. Although bacterial growth was noted only in the flasks which contained 50 and 100 mgL^{-1} EG, the isolate 3a2 was not able to degradeEG. Several studiesreported the toxic effects of EG on microorganisms (Fowles et al.2017; Snellings et al.2013). However, (Chidambara et al.1997) indicated that low concentrationsof EG (13 mgL^{-1}) can beeasily metabolized by a bacterial consortium.

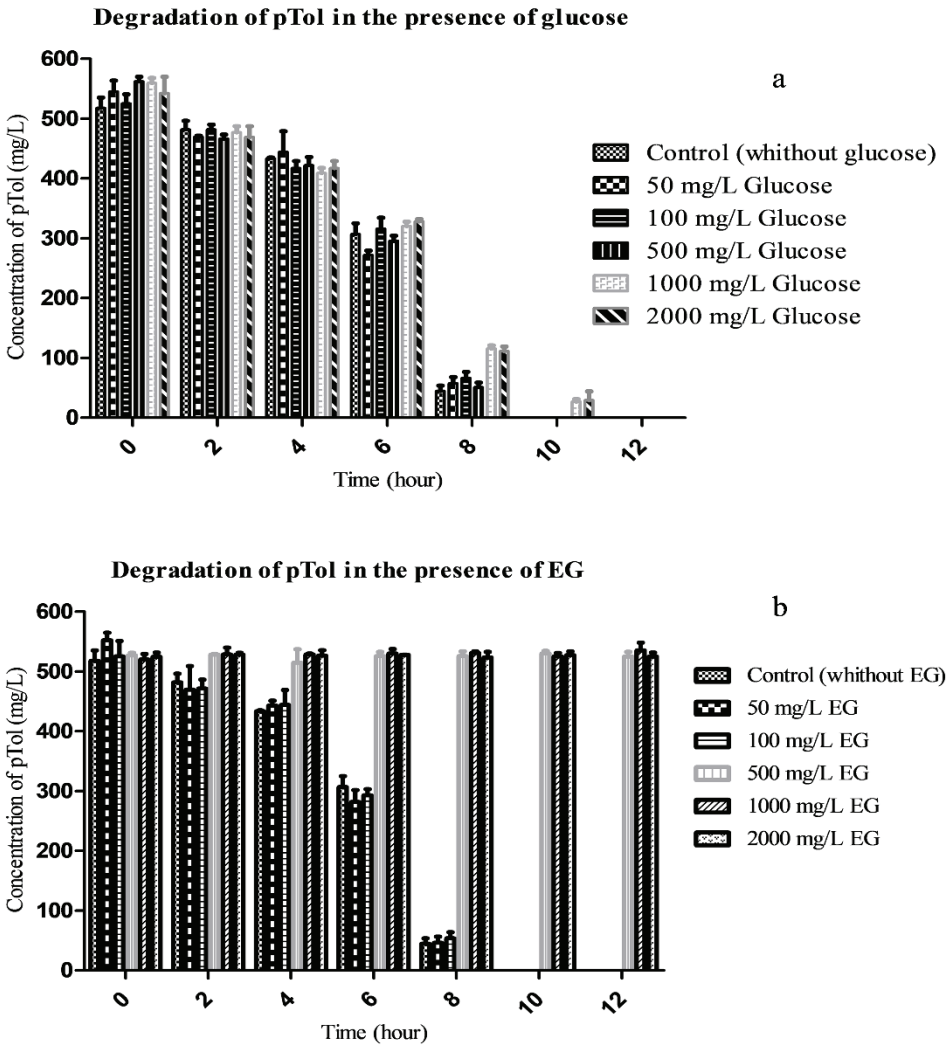

Figure 4. Degradation of pTol in the presence of glucose (a) and EG (b)

Identification and expression of aromatic ring-hydroxylating dioxygenase genes

In this study, PCR amplification allowed the detection of TADO, PDO and TPADO genes in the gDNA of isolate 3a2, but *BADO, ANTDO, TsaM* and *TsaD* geneswere not amplified. The obtained amino acid sequences were compared to reference sequences using the NCBI BLASTP search tool. Phylogenetic trees were constructed with the sequences downloaded from GenBank and UniProt. The amino acid sequence of *TADO, PDO* and *TPADO* genes demonstrated high similarity with the aromatic ring-hydroxylating dioxygenase (RHD) genes of *Comamonas* and *Pseudomonas* strains (Fig. 5).

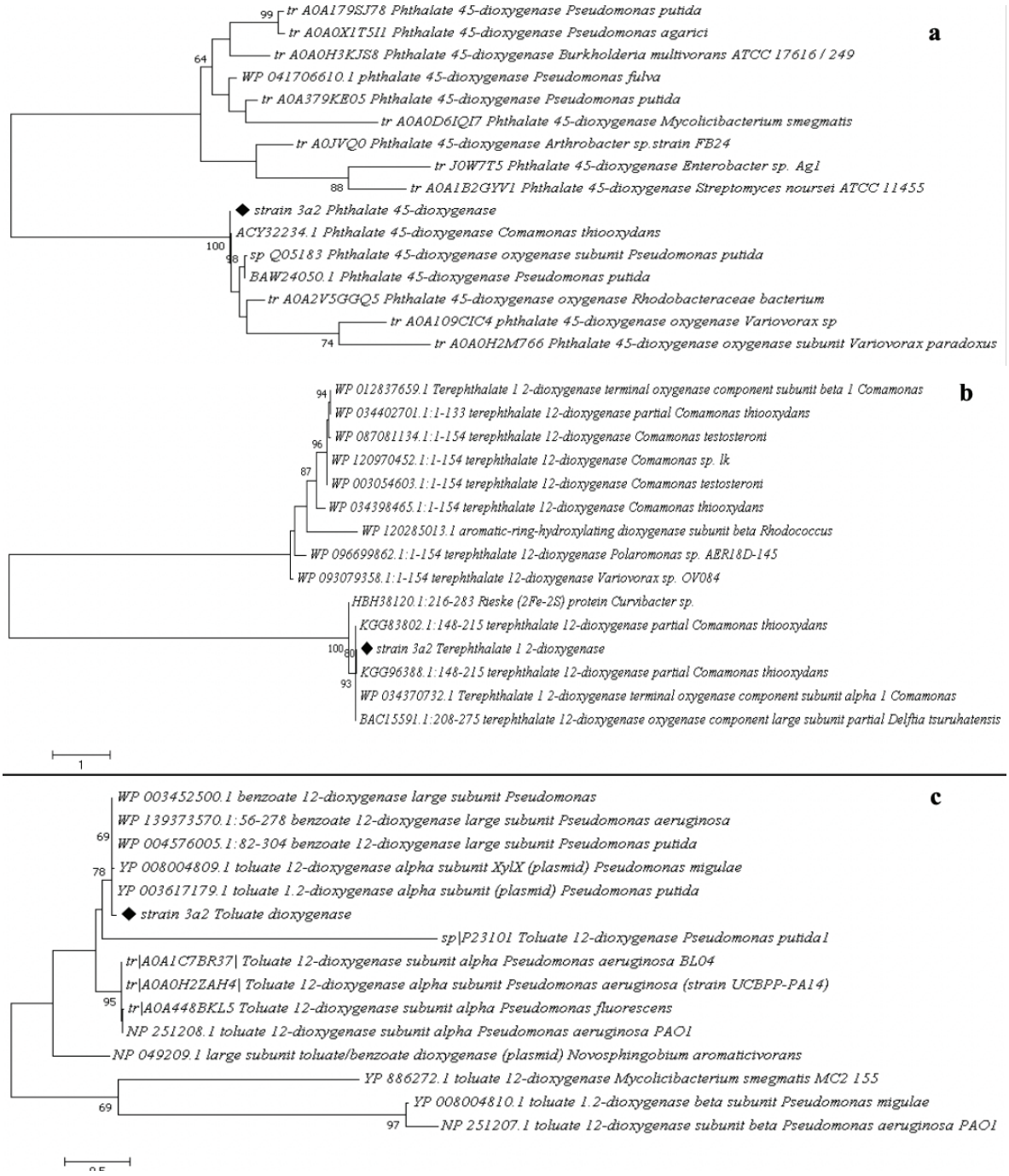

Figure 5. Dendrogram of 3a2 with the aromatic ring-hydroxylating dioxygenase genes. (a) phthalate 4,5 dioxygenase, (b) terephthalate 1,2 dioxygenase, (c) toluate 1,2-dioxygenase

In this study, total RNA was extracted from strain 3a2 during the degradation of pTol at the 2nd, 4th, 6th, 8th and 10th h (Fig.6). The transcript levels of the *TADO, PDO* and *TPADO*genes were detected by real-time PCR. The induction of *TADO* gene started at the beginning of incubation to reach the highest expression at the 6th h (40±3.2 fold). Ge *et al.*(2002) reported thatthe TADO enzyme can catalyze the degradation of wide range of substituted benzoates in the following order of specificity: 3-methylbenzoate > benzoate ≥ 3-chlorobenzoate > 4-methylbenzoate ≥ 4-chlorobenzoate >> 2-methylbenzoate ≥ 2-chlorobenzoate. The induction of *TPADO* and *PDO* genes started at the 4thand 6thhour with induction level equivalent to 18 ± 0.1 and 18.5 ± 0.4 fold respectively.

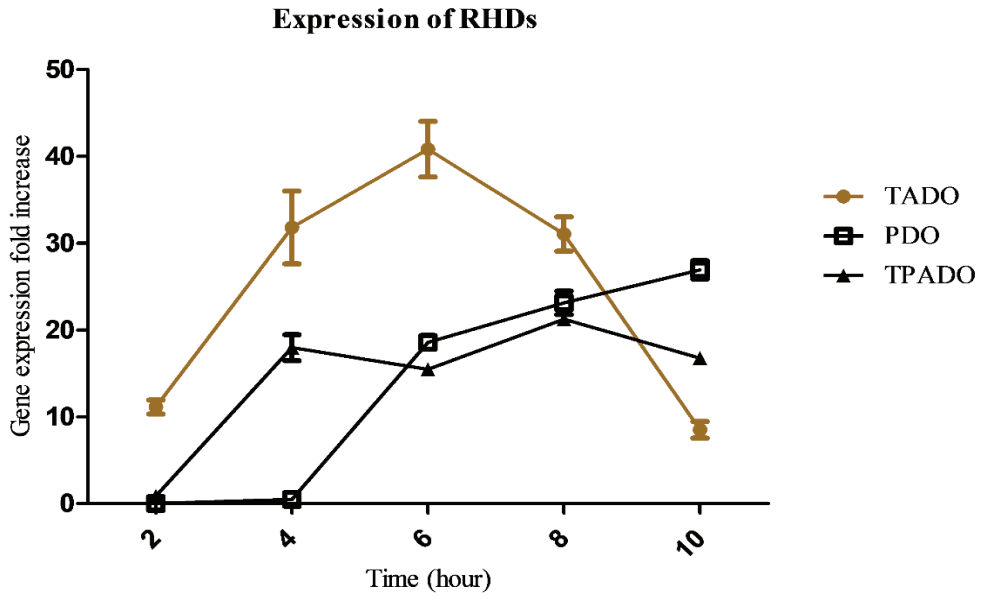

Figure 6. Expression of RHDs

Discusssion

The degradation of aromatic hydrocarbons by *Comamonas testosteroni* strains has been documented (Chen et al.2016; Zhao et al. 2015). Tobajas et al. (2012) reported that *C. testosteroni* can efficiently degrade phenolic compounds. Another study proceeded by Boon et al. (2000) revealed that a *C. testosteroni* isolated from activated sludge was able to use 3-chloroaniline (3-CA) as a carbon source.These results were similar to this study, which reported that a new strain 3a2 *Comamonas testosterone* degrade of pTOL.

PTA is produced by oxidation of p-xylene using the AMOCO process and which results in the production of pTol as an intermediate product (Tomas et al.2013). Also, the ability of bacterial species to convert aerobically pTol to TPA has been reported in previous studies (Junker et al.1997; Locheret al.1991). Luo and Lee

2017 have reported the production of TPA from p-xylene using an engineered *Escherichia coli* system. p-xylene is first converted to pTol (upstream pathway) and thento TPA (downstream pathway)(Luo and Lee 2017).In this study, the formation of phthalic acid (PA) and terephthalic acid during pTol degradation by the isolate 3a2.

Tobajas et al. (2012) reported that glucose does not affect degradation of hydrocarbon. Gao and Skeen (1999)reported that the addition of a carbon source, such as glucose, can induce degradation of chemicals by microorganisms. Another study demonstrated that the biodegradation of 1H-1,2,4-triazole by *Shinella* sp. NJUST26 was enhanced by the addition of low concentration of glucose, sucrose or yeast extract (Wu et al. 2016). Jiang et al. (2019) reported a positive correlation between the growth rate of *Bacillus cereus* strain WD-2 and the biodegradation rate of prochloraz-manganese in the presence of different carbon sources.The effect of different carbon sources on the growth rate and biodegradation rate can be summarized as follows glucose> sucrose> maltose> starch. However, Abdelhafid et al. (2000) demonstrated that the presence of different carbon sources in the same flask might inhibit the biodegradation process.In this study, presence of glucose as carbon source did not influence the pTol degradation by the isolate 3a2.

Bacterial cells are challenged by the availability of nutrients during the biodegradation of hydrocarbons (Obinna et al. 2015). In this study, the effect of organic and inorganic nitrogen sources on pTol degradation by strain 3a2 was evaluated. The results indicated that the biodegradation yield of pTol was higher in the presence of organic nitrogen sources when compared to that recorded in presence of inorganic nitrogen.The effect of organic or inorganic nitrogen sources on the biodegradation of hydrocarbons by microorganisms has been previously studied (Armenante et al. 1995; Abdelhafid et al. 2000; Obinna et al. 2015). Bereka 2013, investigated the effect of different nitrogen sources such as peptone, yeast extract, $(NH_4)_2SO_4$, NH_4Cl_2 and $NaNO_3$ on crude oil biodegradation. The study mentioned that peptone was the optimal nitrogen source as it caused the removal of 73.3 % of crude oil (Bereka 2013). In another study, He et al. (2019) reported that urea is unsuitable for biodegradation of n-hexane by *Pseudomonas* sp. strain NEE2. However, the authors indicated that there was no significant difference of n-hexane removal efficiency in the presence of $(NH_4)_2SO_4$, yeast extract and peptone between 0 to 33h of incubation. Yet long extended incubation periods beyond 33h, the degradation of n-hexane in the presence of $(NH_4)_2SO_4$ was significantly higher than that of organic nitrogen sources (yeast extract and peptone).

Several reviews have covered different enzymes andpathways that are responsible for degradation of aromatic hydrocarbons by various microorganisms.These includeBenzoate-1,2-dioxygenase from *Alcaligenes*

eutrophus B9, toluate-1,2-dioxygenase from *Pseudomonas putida* mt-2, naphthalene 1,2-dioxygenase from *Pseudomonas* sp. strain NCIB 9816, anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1 (Eby et al. 2001; Gao et al. 2010; Joshi and Walia 1996; Vermaet al.2019; Reineke and Knackmuss 1978). Many studies reported the implication of RHD genes in the degradation of aromatic hydrocarbons(He *et al.* 2018; Parales and Resnick 2006; Phale et al.2007; Zhao et al. 2015). Schlafli et al.(1994)reported the ability of *C. testosteroni* T-2 to degrade terephthalate to 1,2-dihydroxy-3,5-cyclohexadiene 1,4-dicarboxylic acid through TPADO enzymes.

TsaMBCD is a metabolic pathway that degrades pTol which eventually results in the formation of TPA. This pathway comprises three stages that are catalyzed by the enzymes TsaMB, TsaC and TsaD during which 4-carboxybenzyl alcohol, 4-carboxybenzaldehyde, and terephthalic acid are respectively formed. TsaMBCDenzymes are encoded by TsaMB, TsaC and TsaD genes respectively in *C. testosteroni* T-2 (Junker et al. 1997; Luo and Lee 2017; Tralau *et al.*2001). Aromatic ring hydroxylating dioxygenase genes are known to catalyze the first step in the biodegradation process of diverse aromatic compounds (Parales and Resnick 2006). Real-time PCR has been used to determine the expression of functional genes of hydrocarbon-degrading bacteria (Beller et al. 2002; Silva and Alvarez 2007). Similarly, Lillis et al. (2010) employed real-time PCR method to assess the abundance and expression of two catechol dioxygenase genes in soil contaminated with 2,4-dichlorophenol over a 21 day biodegradation period (Lillis et al. 2010)

In this study,TPA was produced during the biodegradation of pTol by 3a2 even thoughTsaM and TsaD genes were not detected. Besides, the formation of PA during pTol degradation by 3a2 might have been caused by another pathway. Another study reported the ability of TADO produced by *Pseudomonas putida* mt-2 to catalyzethe dihydroxylation of benzoates (Nakazawa and Yokota 1973). The TADO enzyme is encoded by the xylXYZ genes. These genes are part of the xyl operon which is located on the plasmid pWW0 of *Pseudomonas putida*. The pWW0 plasmid is involved in the biodegradationof xylenes and substituted toluenes (Harayama and Rekik 1990; Kasai et al. 2001).

This study demonstrated the ability of strain 3a2 to degrade pTol at high concentrations upto 1000mgL^{-1}after 10h. The biodegradation of pTol was enhanced in the presence of organic nitrogen compared to the presence of their inorganic counterparts. The isolate was also able to degrade phthalate and terephthalate which resulted from pTol degradation within 24h.Three aromatic ring-hydroxylating dioxygenase genesnamely *TADO*, *TPADO* and *PDO*, which are responsible for the degradation of pTol, TPA and PA respectively, were detected in gDNA of *Comamonas testosteroni*strain 3a2. The high induction of *TADO*, *TPADO* and *PDO* genes indicates the efficient degradation

of hydrocarbons byisolate3a2.According to our knowledge, this is the first study that reports the ability of isolate 3a2 to convert pTol to PA.However, more investigations are required to elucidate the different pathways employed by the strain 3a2 to produce phthalate and terephthalate from pTol.

Acknowledgement

The authors wish to thank Ege University, Scientific Research Projects Fund, (BAP FKB-2019-20395) for the financial support of this study. This work supported by the Ministry of Development Ege University Application and Research Center for Testing and Analysis (EGE-MATAL; 2010K120810). The bacteria used in the study were isolated within the TUBITAK project. The authors wish to thank TUBİTAK (Project number: 112D044-TUBİTAK).

Conflict of Interest: The authors declare that they have no conflict of interest exists.

References

Abdelhafid R, Houot S, Barriuso E (2000) Dependence of atrazine degradation on C and N availability in adapted and non-adapted soils. Soil Biol Biochem 32(3):389-401

Abd-El-Haleem D, Beshay U, Abdelhamid AO, Moawad H, Zaki S (2003) Effects of mixed nitrogen sources on biodegradation of phenol by immobilized *Acinetobacter* sp. strain W-17. Afr J Biotechnol 2(1):8-12

Armenante PM, Fava F, Kafkewitz D (1995) Effect of yeast extract on growth kinetics during aerobic biodegradation of chlorobenzoic acids. Biotechnol Bioeng 47(2):227-33. htttp://doi.org/10.1002/bit.260470214

Beller HR, Kane SR, Legler TC, Alvarez PJJ (2002) A Real-Time polymerase chain reaction method for monitoring anaerobic, hydrocarbon-degrading bacteria based on a catabolic gene. Environ Sci Technol 36:3977-3984

Bereka MM (2013) Towards efficient crude oil degradation by *Pseudomonas* sp. strain-O2: application of plackett-burman design for evaluation of cultivation conditions. Afr J Microbiol Res 7(39):4722-4729

Boon N, Goris J, Paul DV, Paul V, Verstraete W, Top EM (2000) Bioaugmentation of activated sludge by an indigenous 3-chloroaniline- degrading *Comamonas testosteroni* strain,I2gfp. Appl Environ Microbiol 66(7):2906-2913. http://doi.org/10.1128/AEM.66.7.2906-2913.2000.

Chen YL, Wang CH, Yang FC, Ismail W, Wang PH, Shih CJ, Wu YJ, Chiang YR (2016) Identification of *Comamonas testosteroni* as an androgen degrader in sewage. Sci Rep 6:1-13.http://doi.org/10.1038/srep35386.

Chidambara RBB, Ramkumar N, Sıraj AHJ, Chidambaram CJ (1997) Isophthalic, toluic and terephthalic acids using a mixed culture : effluents of pta production. Trans I Chem E 75B:245-256

Eby DM, Beharry ZM, Coulter ED, Kurtz DM (2001) Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp.strain ADP1. J Bacteriol 183(1):109-118.http://doi.org/10.1128/JB.183-1.109–118.2001

Fowles J, Banton M, Klapacz J, Shen H (2017) A toxicological review of the egseries : commonalities and differences in toxicity and modes of action. Toxicol Lett 278:66-83. http://dx.doi.org/10.1016/j.toxlet.2017.06.009

Gao J, Skeen RS (1999) Glucose-induced biodegradation of cis-dichloroethylene under aerobic conditions. Wat Res 33(12):2789-2796

Gao J, Ellis LBM, Wackett LP (2010) The university of minnesota biocatalysis/ biodegradation database: improving public access. Nucleic Acid Res 38:488-491

Ge Y, Eltis LD (2003) Characterization of hybrid toluate and benzoate dioxygenases. J Bacteriol 185(18):5333-5341. http://doi.org/10.1128/JB.185.18.5333-5341.2003

Ge Y, Vaillancourt FH, Agar NYR, Eltis LD (2002) Reactivity of toluate dioxygenase with substituted benzoates and dioxygen. J Bacteriol 184(15):4096-4103. http://doi.org/10.1128/JB.184.15.4096

Guo Q, Zhang J, Wan R, Xie S (2014) Impacts of carbon sources on simazine biodegradation by *Arthrobacter* strain SD3-25 in liquid culture and soil microcosm. Int Biodeter Biodegr 89:1-6. http://doi.org/10.1016/j.ibiod.2013.12.018.

Harayama S, Rekik M (1990) The meta cleavage operon of tol degradative plasmid PWWO Comprises 13 Genes. Molecular & General Genetics 221(1):113-120

He C, Li Y, Huang C, Chen F, Ma Y (2018) Genome sequence and metabolic analysis of a fluoranthene-degrading strain. Front Microbiol 9:1-10. http://doi.org/10.3389/fmicb.2018.02595

He S, Ni Y, Lu L, Chai Q, Liu H, Yang C (2019) Enhanced biodegradation of n-hexane by *Pseudomonas* sp. strain NEE2. Sci Rep 9(1):1-9. https://doi.org/10.1038/s41598-019-52661-0

Heipieper HJ, Martinez HM (2010) Toxicity of Hydrocarbons to Microorganisms. Handbook of Hydrocarbon and Lipid Microbiology 9:1565-1573

Jiang J, Tang M, Chen J, Yang Y (2019)Identification and degradation characteristics of B*acillus cereus* strain WD-2 isolated from prochloraz-manganese-contaminated soils. PLOS ONE14(8):1-13. http://doi.org/10.1371/journal.pone.0220975

Joshi B, Walia S (1996) PCR Amplification of catechol 2,3-dioxygenase gene sequences from naturally occurring hydrocarbon degrading bacteria isolated from petroleum hydrocarbon contaminated groundwater. FEMS Microbiol Ecol19(1):5-15

Junker F, Kiewitz R, Cook AM (1997) Characterization of the p-toluenesulfonate operon tsambcd and tsar in *Comamonas testosteroni* T-2. J Bacteriol 179(3):919-927. http://doi.org/10.1128/jb.179.3.919-927

Kasai Y, Inoue J, Harayama S (2001) The tol plasmid pww0 xyln gene product from *pseudomonas putida* is involved in m-xylene uptake. J Bacteriol 183(22):6662-6666

Kusumawardhani H, Hosseini R, De Winde JH (2018) Solvent tolerance in bacteria : fulfilling the promise of the biotech era ? Trends Biotechnol 36(10):1025-1039. https://doi.org/10.1016/j.tibtech.2018.04.007.

Lillis L, Clipson N, Doyle E (2010) Quantification of catechol dioxygenase gene expression in soil during degradation of 2,4-dichlorophenol. FEMS Micobiol Ecol 73:363-369. http://doi.org/10.1111/j.1574-6941.2010.00906.x

Livak KJ, Schmittgen TD (2001)Analysis of relative gene expression data using real- time quantitative pcr and the 2 ϰ c t method. Methods Elsevier 408:402-408

Locher HH, Leisinger T, Cook AM (1991) 4-Sulphobenzoate 3,4-dioxygenase purification and properties of a desulphonative two-component enzyme system from *Comamonas testosteroni* T-2. Biochem J 274:833-842

Luo ZW, Lee SY (2017) Biotransformation of p-xylene into terephthalic acid by engineered *Escherichia coli*. Nat Commun 8:1-8. http://dx.doi.org/10.1038/ncomms15689.

Maki T, Takeda K (2000) Benzoic acid and derivatives. Ullmann's Encyclopedia of Industrial Chemistry 41-93

Nakazawa T, Yokota T (1973) Benzoate metabolism in *Pseudomonas putida* (arvilla)mt-2:demonstration of two benzoate pathways. J Bacteriol 115:262-267

Noyola A,Macarie H, Varela F, Landrieu S, Marcelo R, Rosas MA (2000) Upgrade of a petrochemical wastewater treatment plant by an upflow anaerobic pond. Water Sci Technol 42(5-6):269-276

Obinna AO, Ikechukwu NEO, Chibuike SU, Alexandra N (2015) Effect of different organic and inorganic nitrogen sources on the kinetics of the breakdown of crude oil using *Pseudomonas*. J Bio Sci 15(1):16-24. http://doi.org/10.3923/jbs.2015.16.24

Parales RE, Resnick SM (2006) Aromatic ring hydroxylating dioxygenases. *Pseudomonas* Chapter 4

Phale PS, Aditya B, Majhi PD, Deveryshetty J, Vamsee-Krishna C, Shrivastava R (2007) Metabolic diversity in bacterial degradation of aromatic compounds. OMICS 11(3):252-279. http://doi.org/10.1089/omi.2007.0004

Pillai CK, Kwon TO, Moon S (2009) Degradation of wastewater from terephthalic acid manufacturing process by ozonation catalyzed with Fe^{+2}, H_2O_2 and UV light: Direct versus indirect ozonation reactions. Appl Catal B 91:319-328. http://doi.org/10.1016/j.apcatb.2009.05.040.

Reineke W, Knackmuss HJ (1978) Chemical structure and biodegradability of halogenated aromatic compounds substituent effects on 1,2-dioxygenation of benzoic acid. Biochimica et Biophysica Acta 542:412-423

Schlafli HR, Weiss MA, Leisinger T, Cook AM (1994) Terephthalate 1,2-dioxygenase system from *comamonas testosteroni* t-2: purification and some properties of the oxygenase component. J Bacteriol 176(21):6644-6652

Shirota M, Seki T, Tago K, Katoh H, Marumo H, Furuya M, Shindo T, Ono H (2008) Screening of toxicological properties of 4-methylbenzoic acid by oral administration to rats. J Toxicol Sci 33(4):431-445. https://doi.org/10.2131/jts.33.431

Silva MLBD, Alvarez PJJ (2007) Assessment of anaerobic benzene degradation potential using 16S rrna gene-targeted real-time PCR. Environ Microbiol 9(1):72-80

Snellings WM, Corley RA, McMartin KE, Kirman CR, Bobst SM (2013) Oral Reference dose for egbased on oxalate crystal-induced renal tubule degeneration as the critical effect. Regul Toxicol Pharmacol 65(2):229-241. http://dx.doi.org/10.1016/j.yrtph.2012.12.005.

Tobajas M, Monsalvo VM, Mohedano AF, Rodriguez JJ (2012) Enhancement of cometabolic biodegradation of 4-chlorophenol induced with phenol and glucose as carbon sources by C*omamonas testosteroni*. J Environ Manag 95:116-121 http://dx.doi.org/10.1016/j.jenvman.2010.09.030.

Tomas RAF, Bordado JCM, Gomes JFP (2013) P-xylene oxidation to terephthalic acid: a literature review oriented toward process optimization and development. Chemical Reviews 113(10):7421-7469

Tralau T, Cook AM (2001) Map of the IncP1b plasmid pTSA Encoding the widespread genes (tsa) for p-toluene sulfonate degradation in C*omamonas testosteroni* T-2. Appl Environ Microbiol 67(4):1508-1516

Varjani SJ, Upasani VN (2016) Carbon spectrum utilization by an indigenous strain of *Pseudomonas aeruginosa* NCIM 5514: production, characterization and surface active properties of biosurfactant. Bioresour Technol 221:510-516

Varjani SJ, Upasani VN (2017) A new look on factors affecting microbial degradation of petroleum hydrocarbon pollutants. Int Biodeter Biodegr 120:71-83

Verma N, Kantiwal U, Yogendra N, Yadav KY, Teli S, Goyal D, Janmejay P (2019) Catalytic promiscuity of aromatic ring-hydroxylating dioxygenases and their role in the plasticity of xenobiotic compound degradation. Microbial Metabolism of Xenobiotic Compounds 10:1-31. http://link.springer.com/10.1007/978-981-13-7462-3.

Wu H, Shen J, Wu R, Sun X, Li J, Han W, Wang L (2016) Biodegradation mechanism of 1H-1,2,4-triazole by a newly isolated strain *Shinella* sp. NJUST26. Sci Rep 6:1-10. http://dx.doi.org/10.1038/srep29675

Xie S, Wan R, Wang Z, Wang Q (2013) Atrazine biodegradation by *Arthrobacter* strain DAT1 : effect of glucose supplementation and change of the soil microbial community. Environ Sci Pollut Res Int 20:4078-4084. http://doi.org/10.1007/s11356-012-1356-4

Zhang YM, Sun YQ, Wang ZJ, Zhang J (2013) Degradation of terephthalic acid by a newly isolated strain of *Arthrobacter* sp. 0574. S Afr J Sci 109(7/8):1-4. http://dx.doi.org/10.1590/sajs.2013/20120019.

Zhang Y, Hong PY, Le Chevallier MW, Liu WT (2015) Phenotypic and phylogenetic identification of coliform bacteria obtained using 12 coliform methods approved by the US Environmental Protection Agency. Appl Environ Microbiol 81:6012-6023

Zhao D, Liu C, Liu L, Zhang Y, Liu Q, Wu WM (2011) Selection of functional consortium for crude oil-contaminated soil remediation. Int Biodeter Biodegr 65:1244-1248. http://dx.doi.org/10.1016/j.ibiod.2011.07.008

Zhao B, Hua X, Wang F, Dong W, Li Z, Yang Y, Cui Z, Wang M (2015) Biodegradation of propyzamide by *Comamonas testosteroni* W1 and cloning of the propyzamide hydrolase gene CamH. Biores Technol 179:144-149. http://dx.doi.org/10.1016/j.biortech.2014.12.009.