REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

DETERMINATION OF ANFIS MEMBERSHIP FUNCTION AND ITS NUMBER BY USING BOX-BEHNKEN EXPERIMENTAL DESIGN METHOD

Esat Sinan ALTIN

MASTER OF SCIENCE THESIS

Department of Chemical Engineering

Chemical Engineering Program

Supervisor

Assoc. Prof. Dr. İlknur KÜÇÜK

Co-Supervisor

Assoc. Prof. Dr. Dilek DURANOĞLU

February, 2023

REPUBLIC OF TURKEY

YILDIZ TECHNICAL UNIVERSITY

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

DETERMINATION OF ANFIS MEMBERSHIP FUNCTION AND ITS NUMBER BY USING BOX-BEHNKEN EXPERIMENTAL DESIGN METHOD

A thesis submitted by Esat Sinan ALTIN in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE is approved by the committee on 22.02.2023 in Department of Chemical Engineering, Chemical Engineering Program.

Assoc. Prof. Dr. İlknur KÜÇÜK
Yıldız Technical University
Supervisor

Assoc. Prof. Dr. Dilek DURANOĞLU
Yıldız Technical University
Co- supervisor

Approved By the Examining Committee

Assoc. Prof. Dr. İlknur KÜÇÜK, Supervisor
Yildiz Technical University _____

Assoc. Prof. Dr. Emel AKYOL, Member
Yildiz Technical University _____

Prof. Dr. İsmail İNCİ, Member
İstanbul University-Cerrahpaşa _____

I hereby declare that I have obtained the required legal permissions during data collection and exploitation procedures, that I have made the in-text citations and cited the references properly, that I haven't falsified and/or fabricated research data and results of the study and that I have abided by the principles of the scientific research and ethics during my Thesis Study under the title of Determination of ANFIS Membership Function and Its Number by Using Box-Behnken Experimental Design Methods supervised by Assoc. Prof. Dr. İlknur KÜÇÜK and Assoc. Prof. Dr. Dilek DURANOĞLU. In the case of a discovery of false statement, I am to acknowledge any legal consequence.

Esat Sinan ALTIN

Dedicated to my wife

and my family

ACKNOWLEDGEMENTS

I would like to thank my mentor dear Assoc. Prof. Dr. İlknur KÜÇÜK for her support, patience and her kind guidance.

I would like to thank my esteemed mentor Assoc. Prof. Dr. Dilek DURANOĞLU, who inspired me, believed in me, and guided me with her vast knowledge and experience, since the first course I took as a chemical engineering student, throughout my undergraduate and graduate education.

In this hard and long path of learning my lovely family showed me great support and understanding. I am grateful for their presence.

My special thanks to my lovely wife, best friend and teacher for her partnership in my all journeys.

TABLE OF CONTENTS

LIST OF SYMBOLS vii
LIST OF ABBREVIATIONS viii
LIST OF FIGURES ix
LIST OF TABLES x
ABSTRACT xi
ÖZET xiii
1 INTRODUCTION 15
1.1 Literature Review 15
1.2 Objective of the Thesis 18
1.3 Hypothesis 18
2 ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM 20
2.1 Fuzzy Logic and ANFIS 20
2.1.1 Fuzzy logic 20
2.1.1.1 Fuzzy logic concept 20
2.1.1.2 Fuzzy clusters and membership functions 21
2.2 ANFIS 21
2.2.1 ANFIS general concept 21
2.2.2 ANFIS structure and layers 21
2.2.3 ANFIS training method 22
2.2.4 ANFIS and membership functions 22
2.2.4.1 Triangular MF(Trimf) 23
2.2.4.2 Trapezoidal MF(Trapmf) 24
2.2.4.3 Generalized bell shape MF(Gbellmf) 24
2.2.4.4 Gaussian MF (Gaussmf) 25
2.3 Data Sorting 25
2.4 Error Calculations and Meaning 26
2.5 ANFIS Results and Analysis 26
3 EXPERIMENTAL DESIGN 28
3.1 Box-Behnken Experimental Design (BBD) 28
3.2 Analysing Results 28
4 MATERIALS AND METHODS 30
4.1 Data Source 30

4.2 ANFIS Application 31
4.2.1 Estimation of adsorption percentage via ANFIS 36
4.2.2 Determination of optimum MF and the number of MF via Box-Behnken experimental design method 37
5 RESULTS AND DISCUSSION **39**
5.1 Determination of MF Types 39
5.2 Determination of Optimum MF Numbers via Experimental Design Approach 40
6 CONCLUSION **45**
REFERENCES **46**
PUBLICATIONS FROM THE THESIS **49**

LIST OF SYMBOLS

wi	Making Estimation Steps
w¯i	Normalizing Outcome Steps
R^2	Mean Squared Error
μAi	Degree of input A
μBi	Degree of input B

LIST OF ABBREVIATIONS

ANFIS	Adaptive Neuro-Fuzzy Inference System
ANOVA	Analysis of Variance
BBD	Box Behnken Design
FIS	Fuzzy Interference System
MFs	Membership Function
pH	Power of Hydrogen
3D	Three Dimensional
IQR	Interquartile Range
RMSE	Root Mean Square Error
Temp	Temperature

LIST OF FIGURES

Figure 2.1 ANFIS architecture 22
Figure 2.2 ANFIS workflow as scheme 23
Figure 2.3 Triangular MF shape 23
Figure 2.4 Trapezoidal MF shape 24
Figure 2.5 Generalized bell shape MF shape 25
Figure 2.6 Gaussian MF shape 25
Figure 2.7 ANFIS rule view interface 27
Figure 4.1 Workspace of MATLAB 31
Figure 4.2 Neuro-Fuzzy Designer of ANFIS 31
Figure 4.3 Adding MF number and types 32
Figure 4.4 Sub. Clustering parameter interface of designer 33
Figure 4.5 General look on Neuro-Fuzzy Designer interface during training 33
Figure 4.6 MF Editor interface of designer 34
Figure 4.7 Closer look at rule section and understanding on selection of MF 35
Figure 4.8 Rule editor of designer 35
Figure 4.9 Surface graph of final model 36

LIST OF TABLES

Table 4.1 Adsorption data 30
Table 4.2 Box-Behnken design matrix 37
Table 4.3 Coded values for Box-Behnken experimental design 38
Table 5.1 RMSE for train and test data obtained at different MF numbers 39
Table 5.2 RMSE obtained from ANFIS via BBD approach 40
Table 5.3 Developed regression models 41
Table 5.4 ANOVA results 42
Table 5.5 ANFIS results and ANOVA prediction comparison of selected MFs 43
Table 5.6 ANFIS prediction results obtained optimum MF and MF numbers 44

ABSTRACT

Determination of ANFIS Membership Function and Its Number by Using Box-Behnken Experimental Design Methods

Esat Sinan ALTIN

Department of Chemical Engineering

Master of Science Thesis

Supervisor: Assoc. Prof. Dr. İlknur KÜÇÜK

Co- supervisor: Assoc. Prof. Dr. Dilek DURANOĞLU

Intelligent algorithms are widely used in the optimization process in chemical engineering Adaptive neuro–fuzzy inference systems (ANFIS) emerged from the combination of fuzzy logic and artificial neural network approaches, and it is a powerful tool to simulate non–linear variation in complex systems.

In this study, adsorption data obtained at different temperature, carbon dosage and pH values were evaluated by using MATLAB ANFIS. In order to achieve prediction via ANFIS with acceptable error values, optimum membership function (MF) and optimum number of MF should be determined. In this study, Box-Behnken experimental design method was used to determine the membership functions (MFs) and the number of MFs systematically. Firstly, the same number of MF were chosen for each input (temperature, carbon dosage and pH), then, by examining the error values, center points and MF types were chosen for the further Box-Behnken experimental design (BBD) evaluations. Constant types of MFs showed acceptable predictions. In order to determine the optimum number of MF for each input, all combinations given in BBD matrix were examined via ANFIS, then, regression

models were developed between the root mean square error and MF numbers for each inputs.

In order to estimate adsorption data via ANFIS, trimf as the optimum MF type; optimum MF numbers for temperature, carbon dosage and pH data were determined as 6, 6 and 3, respectively. The adsorption percentages estimated by ANFIS under optimum conditions were found to be quite close to the experimental values.

Keywords: ANFIS, estimation, membership function, experimental design, optimization

ÖZET

ANFIS Üyelik Fonksiyonları ve Sayılarının Box-Behnken Deneysel Tasarım Yöntemi ile Belirlenmesi

Esat Sinan ALTIN

Kimya Mühendisliği Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Doç. Dr. İlknur KÜÇÜK

Eş-Danışman: Doç. Dr. Dilek DURANOĞLU

Akıllı algoritmalar, kimya mühendisliğinde optimizasyon sürecinde yaygın olarak kullanılmaktadır. Adaptif Ağ Yapısına Dayalı Bulanık Çıkarım Sistemi (ANFIS), bulanık mantık ve yapay sinir ağı yaklaşımlarının birleşiminden ortaya çıkmıştır ve karmaşık sistemlerde doğrusal olmayan varyasyonu simüle etmek için güçlü bir araçtır.

Bu çalışmada farklı sıcaklık, karbon dozajı ve pH değerlerinde elde edilen adsorpsiyon verileri MATLAB ANFIS kullanılarak değerlendirilmiştir. ANFIS ile tahminin kabul edilebilir hata değerleri ile yapılabilmesi için optimum üyelik fonksiyonu (MF) ve optimum MF sayısı belirlenmelidir. Bu çalışmada, üyelik fonksiyonlarını (MF'ler) ve MF'lerin sayısını sistematik olarak belirlemek için Box-Behnken deneysel tasarım yöntemi kullanılmıştır. İlk olarak, her bir girdi (sıcaklık, karbon dozajı ve pH) için aynı sayıda MF seçilmiş, ardından daha sonraki Box-Behnken deneysel tasarım (BBD) değerlendirmeleri için hata değerleri, merkez noktaları ve MF tipleri incelenerek seçilmiştir. Sabit (Constant) tip MF ile kabul

edilebilir tahminler elde edilmiştir. Her girdi için optimum MF sayısını belirlemek amacıyla BBD matrisinde verilen tüm kombinasyonlar ANFIS aracılığıyla incelenmiş ve ortalama hataların karesinin karekökü ile her bir girdinin MF sayısı arasında regresyon modelleri geliştirilmiştir.

Sonuç olarak, adsorpsiyon verilerini ANFIS ile en iyi şekilde modellemek için optimum MF tipi olarak trimf; sıcaklık, karbon dozajı ve pH verileri için optimum MF sayıları sırasıyla 6, 6 ve 3 olarak belirlenmiştir. ANFIS ile optimum koşullarla tahmin edilen adsorpsiyon yüzdeleri deneysel değerlere oldukça yakın bulunmuştur.

Anahtar Kelimeler: ANFIS, tahmin, üyelik fonksiyonu, deneysel tasarım, optimizasyon

1
INTRODUCTION

1.1 Literature Review

Adaptive Neuro-Fuzzy Inference System (ANFIS) is used in different estimation studies, such as reactor productions, battery life, drying of products, wastewater treatment etc. (Sreekumar et al.,2020; Fan et al.,2018; Hong et al.,2018; Teimouri&Sohrabpoor,2013; Dutta&Kumar,2018; Tanhaei et al.,2016). This different estimation studies based on observing the soft computing process of ANFIS. There has been a lot of research done on ANFIS and various approaches to its application.

Sreekumar et al. (2020) studied on "Adaptive neuro-fuzzy approach to sodium chlorate cell modelling to predict cell pH for energy-efficient chlorate production". Sodium production is highly energy intensive electro chemical process. Prediction of pH of the electrolyte cell is main part of efficiency of process. In their study, pH value of electrolyte cell was predicted by flow rates of brine, HCl, NaOH, cell electrolyte temperature, DC load current and pH of feed. ANFIS modelling started with sorting raw data. Data sets were used from 2017 and 2018 resulting 9000 data samples. Principal component analysis and interquartile range was used to remove noise data. In this study, ANFIS with statistical methods and hybrid learning model resulted with prediction accuracy of %88. In their study, ANFIS can predict pH value of cell electrolyte which can be applied to adjusting efficiency of electrolyte cell systems (Sreekumar et al.,2020).

Teimouri and Sohrabpoor (2013) studied "Application of adaptive neuro-fuzzy inference system and cuckoo optimization algorithm for analysing electrochemical machining process". Electrochemical machining process offers special privileges. In their work, electrolyte concentration, electrolyte flow rate, applied voltage and feed rate were used to train ANFIS. Cemented tungsten carbide is a hard and tough composite, electrochemical machining process is necessary for this compound. Parameters were coded as -2, -1, 0, 1 and 2, then ANFIS modelling was carried out

with coded training data set. The aim of this study was to make parameters easy to read and compute by coding each parameter. ANFIS uses 5 layers of computation segments and various function models. Optimization of process by cuckoo algorithm was also used. Triangular membership function (MF) with 2-2-2-2 structure found to be the most precise model for material removal rate but for surface roughness gaussian model was the most precise (Teimouri&Sohrabpoor,2013).

Fan B. et al. (2018) studied on "An Adaptive Neuro-Fuzzy Inference System (ANFIS) Based Model for the Temperature Prediction of Lithium-Ion Power Batteries". Lithium-ion batteries are widely applied in personal electronic devices due to energy/power density, low self-discharge rate and long cycle life. Thermal management is crucial for battery life and performance. In their study ambient temperature, state of charge and current were used as input, cell temperature was used as output. Outward and inward prediction was used. Outward prediction is when training data are selected inside a group and predicting and outside data; inward prediction is when training data are selected around the target data. As a result of their study, gaussian MF and linear model was used and model was achieved with 4.5% error. Their outcome of this study was that ANFIS was a suitable soft computation method for prediction of cell temperature (Fan et al.,2018).

Hong E. et al. (2018) studied "ANFIS based modelling of dewatering performance and polymer dose optimization in a wastewater treatment plant". In their study, biosolid quality to total solid, volatile solid content, volatile fatty acid of digested sludge and polymer dose were used as inputs. Training data were collected for four years. Values were normalized and a total of 500 data were used. Normalizing a value is subtracting minimum value from actual value and dividing by minimum value subtracted from maximum value. Computation of changes inside the wastewater treatment process and predicting outcome can increase efficiency. In this study it was showed that soft computation can accomplish to predict big processes instantaneous values and achieve best outcome by making necessary changes (Hong et al.,2018).

Dutta P. and Kumar A. (2018) studied on "Application of an ANFIS model to optimize the liquid flow rate of a process control system". Input variables drives output variables in liquid processes, in order to predict outcome, mathematical methods

were used. Instead of time consuming and inflexible classical methods ANFIS method was tried because of flexibility and low computing need. 134 data were used to train ANFIS model. Gaussian, triangular and generalized bell shape MF were modelled with 3-3-3, 3-3-3-3, 5-3-3-3 and 3-3-3-3-3 structures. As a result of their study, increasing MF doesn't increase model performance, increasing epoch and training data increase model performance. Their study showed that ANFIS structure is special for the process and different approaches should be applied to obtain better results (Dutta&Kumar,2018).

Bahareh Tanhaei (2016) studied on "Neuro-fuzzy modelling to adsorptive performance of magnetic chitosan nanocomposite". Magnetic chitosan adsorption performance for the methyl orange removal modelled by ANFIS in this study. The pH, adsorbent dosage, initial concentration and contact time were determined as main parameters for this study. As output methyl orange removal percentage rate and methyl orange uptake at contact time was stated and calculation of output data was estimated. This study achieved R^2 >0.99 and stated fundamental parameters for adsorptive performance prediction (Tanhaei et al.,2016).

Using raw data is the most direct way to use ANFIS as this way, data is separated for training and testing data to model process, and the model with least error is selected as final model (Sarkar et al.,2014).

Using huge data sets increases computation work on ANFIS but correct model must be achieved by manipulating MFs, using huge data sets feeds model with more examples so the model can be easily optimized by training procedure (Onu et al.,2021).

Using modelling functions on data sets beforehand increases the success of the model. In order to use correct function, the user must be able to understand the problem and know how to approach to a system with related functions. ANFIS is mostly used for unknown problems to find new solutions so using already known equations to make model is a successful but time and knowledge consuming procedure, which is not ANFIS's main target (Wong et al.,2020).

As observed from the studies found in literature, ANFIS modelling does not depend on a system. In the most studies, there is a lack of understanding and explaining how to model ANFIS and how to determine number of MF and MF structure of ANFIS

model. Studies showed that ANFIS is a practical tool to model equations and get legit results. ANFIS is used with trial-and-error method which causes non-systematic approach with just estimation of the best result. Step-by-step use of ANFIS and understanding each step is crucial to achieve optimum model for each data set. Most studies do not have a system to solve modelling problems but use a certain way to achieve optimum model for current knowledge. This study states a systematic approach to optimize MF and the number of MF to make estimation for adsorption data set via experimental design.

1.2 Objective of the Thesis

Chemical engineering problems include many parameters, and finding relationships among those parameters can be challenging depending on the related system. ANFIS's fast, legitimate estimation and discovery of parameter relationships allows it to make faster progress with unknown systems. Until now, ANFIS application on a data set has relied on trial-and-error or the user's knowledge of the relationship of inputs.

In order to predict any data via ANFIS, the MF types and the number of MF for each input have to be decided before prediction. There is limited information in the literature about selecting the appropriate MFs and the number of them. To use ANFIS for properly sorting data, using the optimum MFs and the number of MFs are crucial. This study aimed to find a systematic way to find the optimum MF type and its number for ANFIS estimation. For that reason, Box-Behnken experimental design method was applied.

1.3 Hypothesis

The approach of this study is to find the best possible way to model ANFIS with a data set without knowing the data set's source. Uncertainty about the source of a data set is an excellent way to use ANFIS because only a universal approach can transform an unknown data set into a meaningful model upon which to make estimations. ANFIS is a program consisting of data, MFs, number of MFs and iterations of the training model. Using the correct approach to a data set with optimized MF and number of MF can achieve better results with less knowledge and less computer power. In order to achieve prediction via ANFIS with acceptable error

values, optimum membership function (MF) and optimum number of MF should be determined. Box-Behnken experimental design (BBD) method can be used to determine the number of MFs systematically. For that reason, BBD matrix can be applied, and a regression model can be developed between the root mean square error obtained from ANFIS and MF number of each input, hence, optimum number of MF for each inputs can be determined. Consequently, using optimum MF numbers, ANFIS estimation can be done with more precise.

2
ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

2.1 Fuzzy Logic and ANFIS

ANFIS is an artificial neural network based on the Takagi-Sugano fuzzy inference system. Technique developed in the early 1990s. The system uses if-then rules to predict the outcome; inputs are pre-processed in the first layer, then processed again in the fuzzification layer. Rules are applied to inputs according to MFs. The computation is carried out depending on the related rule (Wikimedia Foundation,2022).

ANFIS sorts all data and then uses those data in the correct equations to generate estimations with less error than classical calculations. Instead of using complicated equations and computer power, the ANFIS method is easier to use and takes less time to achieve results. There are different ways to provide data to the system, such as by coding, making pre-computations on the data, using raw data, etc. Determining which parameters are important for processing and sorting data according to this information is crucial. After the determination of data, unreliable data is removed with different statistical rules to lessen experimental errors (Sreekumar et al.,2020). When data is provided, MF preparation is started. In this process, gaussian, bell-shaped, triangular, and trapezoidal MF are used. MFs are sorting individual data variables. Rules are applied on inputs according to the MF created and using 3 MF to sort the data means dividing the data into 3 groups: low, medium, and high. When a data set is given to get an estimation, first all the data sets are sorted in the ANFIS structure according to their MF group, then ANFIS uses MF rules to make computations (Talpur et al.,2017).

2.1.1 Fuzzy logic

2.1.1.1 Fuzzy logic concept

Fuzzy logic was introduced by Lotfi A. Zadeh in 1965. Fuzzy logic is a way to express computations according to the clusters used to sort data. A data point can be

classified as [0, 1] if it belongs to a cluster or not. For fuzzy logic, a data can be stated between 0 and 1 which creates infinite outcomes.
All data are divided according to clusters determined by the user, and calculations are done depending on which cluster suits our data better (Talpur et al.,2017).

2.1.1.2 Fuzzy clusters and membership functions

Fuzzy clusters can be determined as normal clusters with intersections. Fuzzy clusters define partial truth depending on the MF used. The advantage of using fuzzy clusters is to determine which calculation is more suitable for the desired outcome. For each inlet, a MF is determined, and these functions are adjusted with training data. Each MF has a unique curve, this way, depending on training data, a more suitable function can be chosen (Talpur et al.,2017).

2.2 ANFIS

2.2.1 ANFIS general concept

ANFIS was introduced by Jang in 1993. The aim of using ANFIS is to take crisp inputs and compute according to the MF to estimate an outcome. The computations are done by several layers inside the ANFIS structure. First-level MFs are used to determine which cluster is more suitable. Following this, IF-THEN rules are applied to the data. Rules are generated from crisp data fed to the system for training. Calculations are carried out according to MFs and rules, and then an estimation of the system can be achieved (Talpur et al.,2017).

2.2.2 ANFIS structure and layers

ANFIS consists of 5 layers, as shown in Figure 2.1 The first layer is for MFs that determine the degree of truth for each input according to MF (μAi, μBi). The second and third layers are for the IF (antecedent) and THEN (consequent) parts of the rule. The second layer estimates (wi) using IF-THEN rules, and the third layer normalizes the result ($\bar{w}i$). The fourth layer is a linear polynomial equation. The fifth layer is simply the sum of the outputs of the previous layer's rules (Talpur et al.,2017).

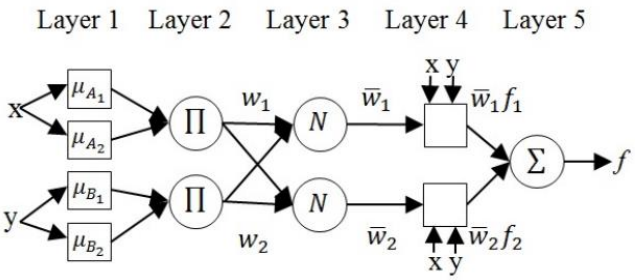


Figure 2.1 ANFIS architecture (Talpur et al.,2017)

2.2.3 ANFIS training method

ANFIS makes estimations with tuneable MF parameters and consequent parameters of a linear polynomial equation on the fourth layer. Training adjusts the first and fourth layers as shown in Figure 2.1.

Training methods tune calculation variables and reshape MFs in order to achieve better clusters to represent inputs and outputs. The training process is considered finished when errors are no longer decreasing (Wu,2012).

Having multiple data sets and performing more iterations increases the likelihood of better training results. The model is shaped and the error value is fixed after a certain amount of training.

Training also determines the limits of the model. The cluster range is set by using limit values, and thus the outcome is also affected. The results page of ANFIS shows possible outputs that can be out of range, but this is only applicable for outputs. The inputs must be within the cluster range.

2.2.4 ANFIS and membership functions

Determining the MF is crucial for training the ANFIS. The ANFIS model is shaped around the data set and how well the data set is divided into clusters. Each set of data has a unique approach to results, as known by general equations and computations. If-then rules of fuzzy logic is only applicable if data clusters are well arranged. There are various MFs but in this study, triangular, trapezoidal, gaussian and generalized bell-shaped functions are focused since they are fundamental and applicable to almost all data sets (Teimouri&Sohrabpoor,2013).

The estimation desired is calculated in the ANFIS structure. Examine the MF distribution to see which inputs will be calculated in which rule. MF graphs cover

the maximum to minimum value of your input, which then determines which MF is more suitable for selected data (Sarkar et al.,2014).

As shown in Figure 2.2 workflow of ANFIS starts with input. Selecting a related fuzzy rule to make calculations. Determining which rules give the best results and finally creating an output (Teimouri&Sohrabpoor,2013).

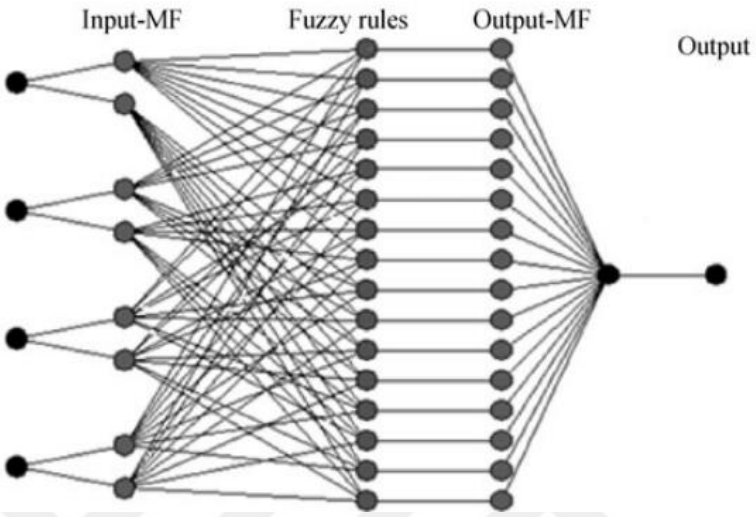


Figure 2.2 ANFIS workflow as scheme (Teimouri&Sohrabpoor,2013)

2.2.4.1 Triangular MF(Trimf)

The triangular MF is the simplest shape among other MFs. As shown in equation 2.1, the equation for triangular MF is determined by three parameters: two for the feet (a and c) of the curve and one for the tip (b) (Talpur et al.,2017).

$$traingular(x, a, b, c) = max\left(min\left(\frac{x-a}{b-a}, 1, \frac{c-x}{c-b}\right), 0\right) \quad (2.1)$$

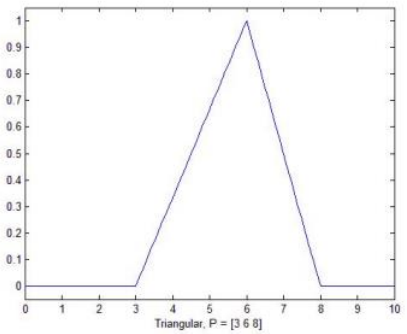


Figure 2.3 Triangular MF shape (Talpur et al.,2017)

The triangular shape creates more specific clusters because of the narrow shape of the MF. Accurate calculations are achieved by using the points closer to the tip point of the MF. The triangular MF is a special kind of trapezoidal MF as it has two tips that collide and form one, as shown in Figure 2.3 (Wu,2012).

2.2.4.2 Trapezoidal MF(Trapmf)

The trapezoidal MF is more like a triangular MF with two tip parameters. As shown in equation 2.2, a trapezoidal MF has two feet as a and d parameters and two tips as b and c parameters. (Talpur et al.,2017).

$$trapzoidal(x, a, b, c, d) = max\left(min\left(\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}\right), 0\right) \tag{2.2}$$

The trapezoidal MFs create wider clusters than triangular MFs as shown on Figure 2.4.

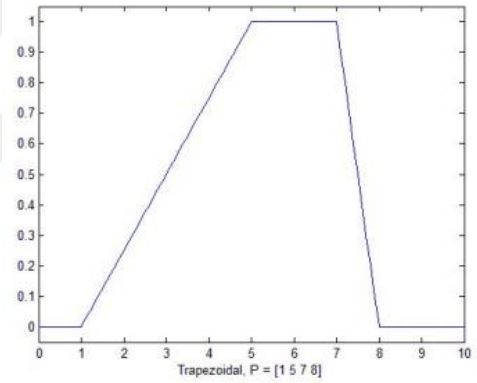


Figure 2.4 Trapezoidal MF shape (Talpur et al.,2017)

2.2.4.3 Generalized bell shape MF(Gbellmf)

The generalized bell shape MF is used to determine a more soft transition between data points, as shown on Figure 2.5. The equation for a generalized bell shape consists of three parameters, which are the width of the curve as parameter a, b as a positive integer, and c as the center of the curve as shown in equation 2.3 (Talpur et al.,2017).

$$bell(x, a, b, c) = \frac{1}{1 + \left|\frac{x-c}{a}\right|^{2b}} \tag{2.3}$$

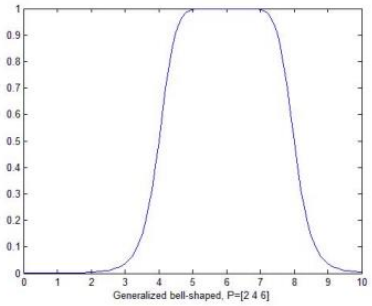


Figure 2.5 Generalized bell shape MF shape (Talpur et al.,2017)

2.2.4.4 Gaussian MF (Gaussmf)

The Gaussian MF creates a curve similar to the generalized bell shape MF, but it only has two parameters: c as the center of the curve, as shown in equation 2.4, and for determining the width of the curve, as shown in Figure 2.6. (Talpur et al.,2017).

$$gaussian(x, c, \sigma) = e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \quad (2.4)$$

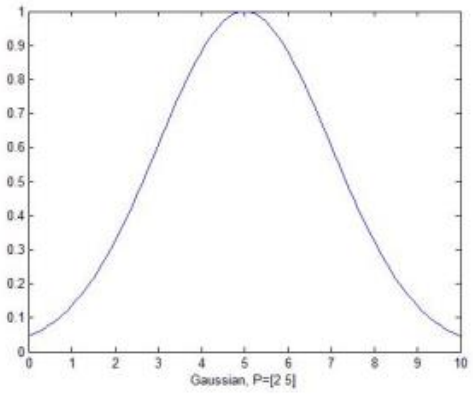


Figure 2.6 Gaussian MF shape (Talpur et al.,2017)

2.3 Data Sorting

Data is one of the key elements for the ANFIS modelling process. Reliable data is needed for this procedure, and different types of statistical approaches are used to achieve a better model. Experimental errors may occur because pre-treatment is needed on data sets.

The interquartile range (IQR) is one of the statistical approaches. With IQR, swings in raw data can be cancelled out, and a smoother data set is achieved (Sreekumar et al.,2020).

Principal component analysis (PCA) is another way to data selection. A graph is used to lay out all of the data in this method, and a line is drawn. Then, a perpendicular line is drawn to see the relationship, and finally, a line is generated that is general for all data and provides a rough estimate to understand mean data sets (Goodarzi & Freitas, 2010).

Gaussian process emulator, this is a statistical model which is expensive and takes time to solve but eventually provides results like ANFIS (Kanwal et al.,2018).

2.4 Error Calculations and Meaning

ANFIS uses RMSE (root mean square error) to calculate the error of the model. All samples are separately calculated and shown in the graph as sample number and output value.

The results page has a calculation section that enables using different inputs to compute and achieve output.

Results are shown as a surface, which has two input axis and one output axis. Surface is used to understand the effect of two inputs by picking other inputs as constants. The surface graph gives the overall results of the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \quad 2.5$$

2.5 ANFIS Results and Analysis

ANFIS calculations are done within the model system. Each fuzzy rule is tested, and the results are all summed up according to the degree of truth to achieve an output. The system does this automatically, but all rules can be adjusted, and the effect of rules can be seen in Figure 2.7.

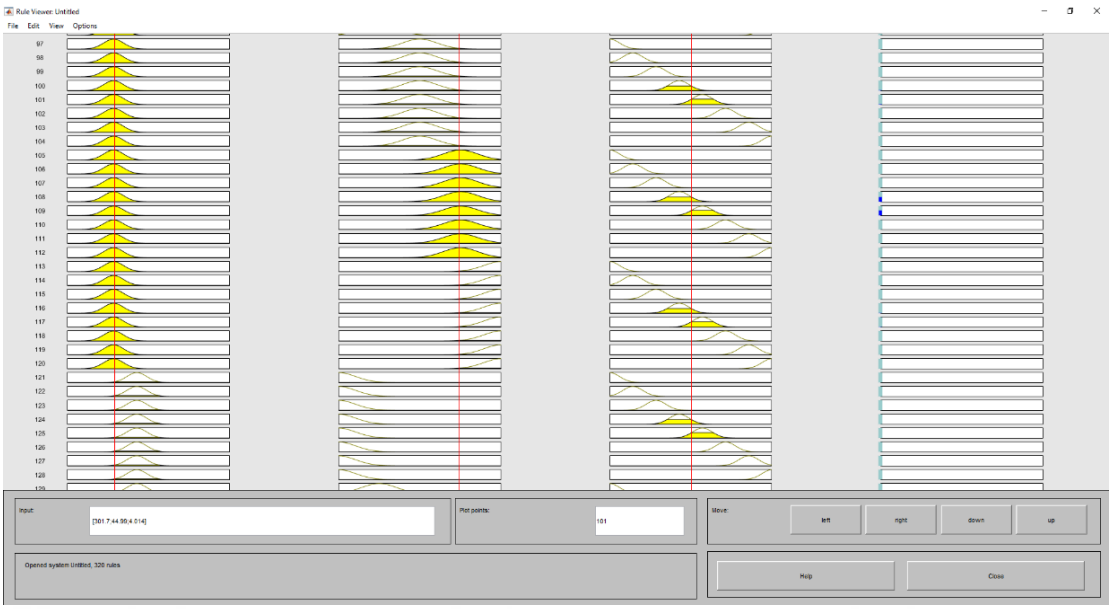


Figure 2.7 ANFIS rule view interface

3
EXPERIMENTAL DESIGN

Experimental design is a method to determine how to approach an experiment. There are several ways to determine how to make an approach. Experimental design aims to achieve final data with the least number of experiments. In order to achieve a minimum amount of experiments, designers should determine parameters, understand the effects of those parameters, and then make an approach. The designer should understand if the design is at fault depending on results and cancel out unwanted data sets to make a clean approach (Rust et al., 2021).

Experimental design is a statistical concept. The main aim is to manage parameters and create an experimental plan. An experimental plan is crucial to achieving a data set with minimum error. A regression model is developed after obtaining the results of the developed experimental plan. The developed model is statistically analysed via ANOVA. ANOVA prepares a report where the designer can analyse which parameter to use for modelling (Rust et al.,2021; Soltany&Zahiri,2019).

3.1 Box-Behnken Experimental Design (BBD)

Box-Behnken is the experimental design method used to determine the best approach for a data set. The Box-Behnken method is applied to a data set by choosing parameters. Parameters are coded as -1,0 and 1 to make needed calculations for modelling. BBD requires 15 experimental runs, three of which are central points. After 15 experiments are carried out, errors are compared, and the best modelling option is shown as a number between -1 and 1. This number can be translated to real numbers, as done in the first stage of the Box-Behnken experimental design (Aslan&Cebeci,2007; Wang et al.,2022).

3.2 Analysing Results

ANOVA results are created as a report. This report contains information on which combination is best among all trials and calculates the errors of combinations of experiments. These results are calculated according to the relevance of parameters.

After achieving the best relevance, the experimental design is completed. Following that, the application is completed, and the results are compared (Rust et al.,2021; Soltany&Zahiri,2019; Aslan&Cebeci,2007; Khoshvaght et al.,2021).

4
MATERIALS AND METHODS

4.1 Data Source

Adsorption data were taken from a previous study (Duranoglu et. al, 2010). Adsorption data obtained at different pH, carbon dosage (mg) and temperature can be seen in Table 4.1. The first 40 data points were investigated for training purposes, while the remaining data points were used as test data in the ANFIS program.

Table 4.1 Adsorption data

No	Temp. (°C) (Input 1)	Carbon dosage (mg) (Input 2)	pH (Input 3)	Adsorption (%)	No	Temp. (°C) (Input 1)	Carbon dosage (mg) (Input 2)	pH (Input 3)	Adsorption (%)
1	295	2.2	2	20.05	**26**	295	8.4	6	25.20
2	295	5.7	2	47.05	**27**	295	20.8	6	56.67
3	295	11.3	2	85.41	**28**	295	30	6	73.37
4	295	13.5	2	94.49	**29**	295	43.1	6	87.48
5	295	20.3	2	99.68	**30**	295	59.8	6	95.55
6	295	31.6	2	99.97	**31**	305	4.9	2	34.90
7	295	40.6	2	99.99	**32**	305	7.8	2	55.41
8	295	59.5	2	99.99	**33**	305	15.4	2	90.79
9	295	2.8	3	22.39	**34**	305	19.8	2	98.12
10	295	5.1	3	37.34	**35**	305	29.2	2	99.72
11	295	7.9	3	54.34	**36**	318	5	2	44.05
12	295	13.9	3	92.95	**37**	318	7.8	2	61.57
13	295	17.7	3	97.79	**38**	318	15	2	95.60
14	295	21.5	3	99.35	**39**	318	20.1	2	99.72
15	295	40.3	3	99.77	**40**	318	30.9	2	99.97
16	295	59.8	3	99.83	**41**	295	7.8	2	63.95
17	295	2.4	4	15.81	**42**	295	17.5	2	98.98
18	295	4.5	4	28.66	**43**	295	10.3	3	67.73
19	295	10.3	4	62.40	**44**	295	29.7	3	99.89
20	295	12.8	4	73.43	**45**	295	8.7	4	53.78
21	295	17.7	4	89.82	**46**	295	30.8	4	99.64
22	295	20.7	4	94.76	**47**	295	11.2	6	32.43
23	295	41.6	4	99.87	**48**	305	10.5	2	67.61
24	295	2.8	6	8.77	**49**	318	10.4	2	78.22
25	295	5.1	6	15.39					

4.2 ANFIS Application

ANFIS is a MATLAB application to make estimations depending on functions prepared by training fuzzy logic with a data set. Train and test data sets can be loaded from a computer or workspace (Fig.4.1). ANFIS has an interface that shows loaded data, generate fuzzy interference system (FIS), train FIS and test FIS as shown in Figure 4.2.



Figure 4.1 Workspace of MATLAB

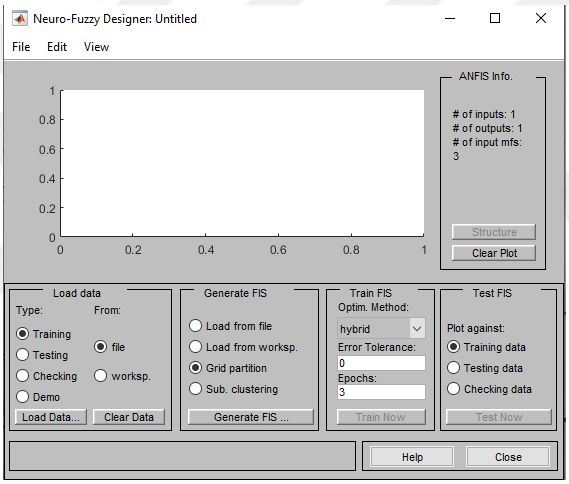


Figure 4.2 Neuro-Fuzzy Designer of ANFIS

The load data section has training, testing, checking, and demo sections. Training data upload is only used by the Train FIS section. Testing and checking data set is only used after training and only used at test FIS section. The Generate FIS section has load from file, load from workspace, grid partition and sub. clustering sections. Loading from workspace and file is only used if the user has a ready to use FIS.

Grid partition is a training method in which the user selects the MF's function, linear or constant application, and the number of MF for each parameter using the interface shown in Figure 4.3. This method is generally used in all studies because the model can be shaped by the user depending on the experimental data. This method initially prepares MF evenly distributed in each parameter but after training

MF distributed depending on their effect on results. The user must determine MF parameters correctly to achieve realistic results (Wu,2012).

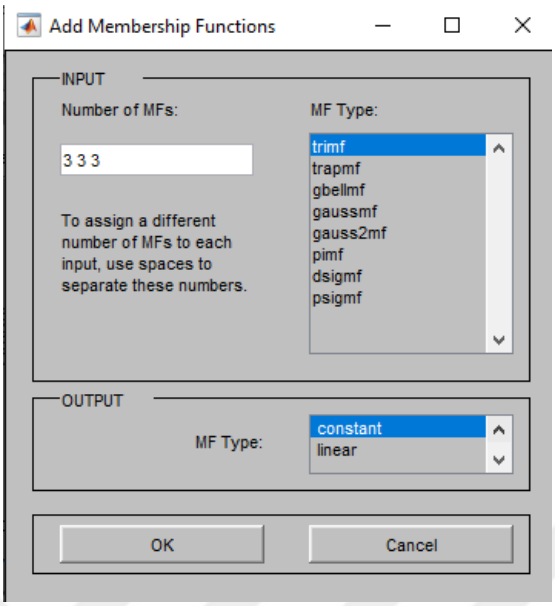


Figure 4.3 Adding MF number and types

Sub clustering is a training method that consists of choosing the final error limit and the general error application as shown in Figure 4.4. This method selects the MF and number of MF based on the specified error value. This method can create a huge number of MF and can be useless for narrow data sets. This method focuses on the raw data set, thus training will not have an effect after the error point is achieved.

The range of the influence value should be in the range of 0 and 1. A smaller range of influence usually creates more and smaller data clusters and produces more fuzzy rules.

The squash factor contains a positive scalar. The smaller the amount of a squash, the more it usually creates more and smaller data clusters.

The value of the acceptance ratio should be in the range of 0 and 1. The acceptance ratio must be greater than the rejection ratio, and the value of the rejection ratio should be in the range of 0 and 1 [1].

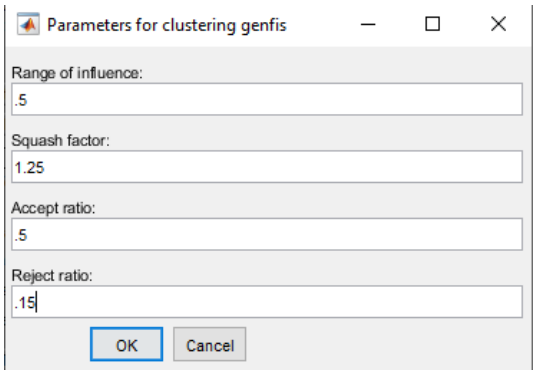


Figure 4.4 Sub. Clustering parameter interface of designer

Training carries on until the error value approaches a constant value. Error calculations are hybrid and backprop. A general look at the Neuro-Fuzzy Designer interface during training can be seen in Figure 4.5.

The training interface of ANFIS shows the error value and epoch number on a graph. The current error value and epoch number can be seen on the left-bottom corner of the interface.

The number of inputs, number of outputs, and number of MF can be seen in the ANFIS info section on the top-right.

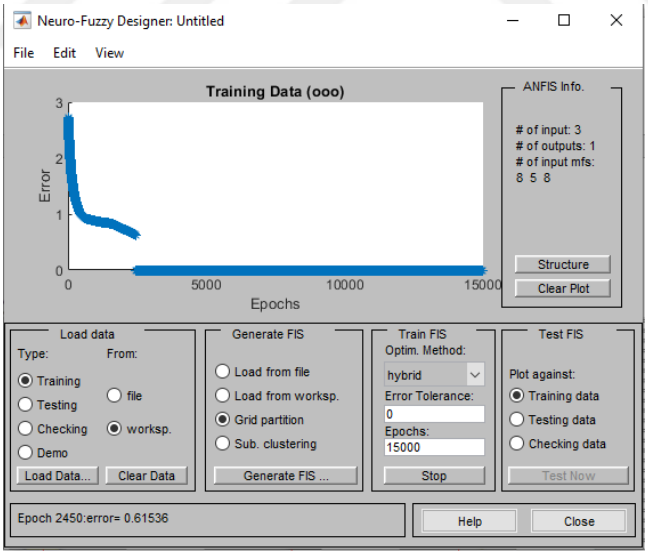


Figure 4.5 General look on Neuro-Fuzzy Designer interface during training

Finally, the resulted model is plotted against training, checking, or testing data, as shown in Figure 4.5. At this point, it is possible to determine which experimental data caused an error and which data was correctly estimated.

MF can be adjusted from the edit page of the ANFIS interface shown in Figure 4.6. This page displays all parameters from the training step that have had their MF adjusted. MF determine a data range, and if the inlet covers most of the area of one MF, that equation is used for calculation. Adjusting MF can directly affect the results. MF can be adjusted by left bottom, tip and right bottom points according to related function parameters and function of the MF.

MF stands for a cluster in a selected data set. This means that a data set is divided by the number of MF. When an input enters the system, all rules are tried by the system, and finally the MF with the most surface covered by the input has more effect on the output calculation.

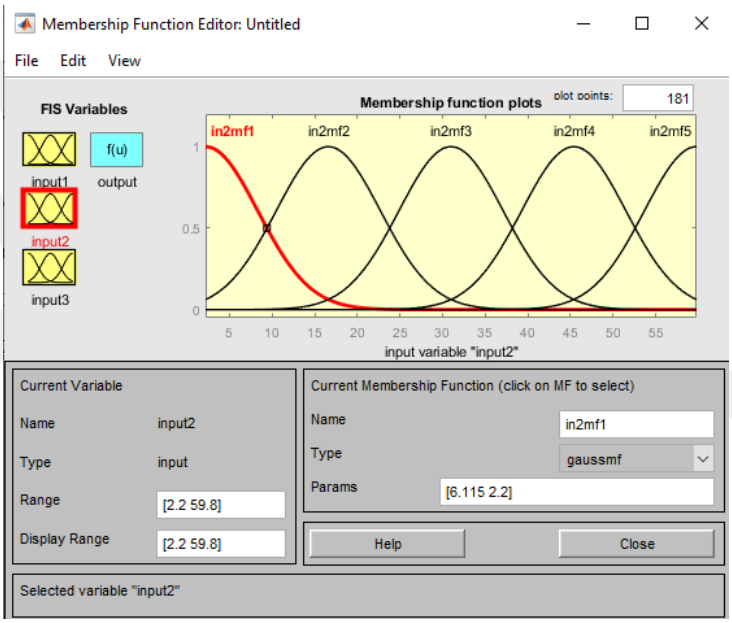


Figure 4.6 MF Editor interface of designer

As seen on Figure 4.7, when an input is entered in ANFIS, all MF are intersected by that input value. Then the intersection gives us an area on the MF. ANFIS selects the biggest area of all inputs and uses related rules to make an estimation.

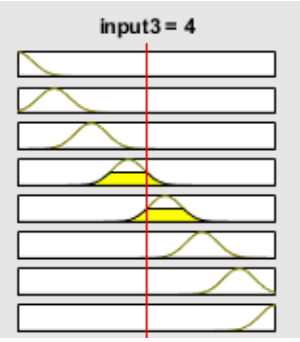


Figure 4.7 Closer look at rule section and understanding on selection of MF

During the modelling procedure, all rules of ANFIS can be changed and created, as shown in Figure 4.8. All clusters are computed in all combinations according to the fuzzy rules, which create rules that include different calculations. In this section, all rules can be deleted, added, or changed depending on the situation. Rules can be created by if-then rules, but connection to MF can be delivered by "and" or "or".

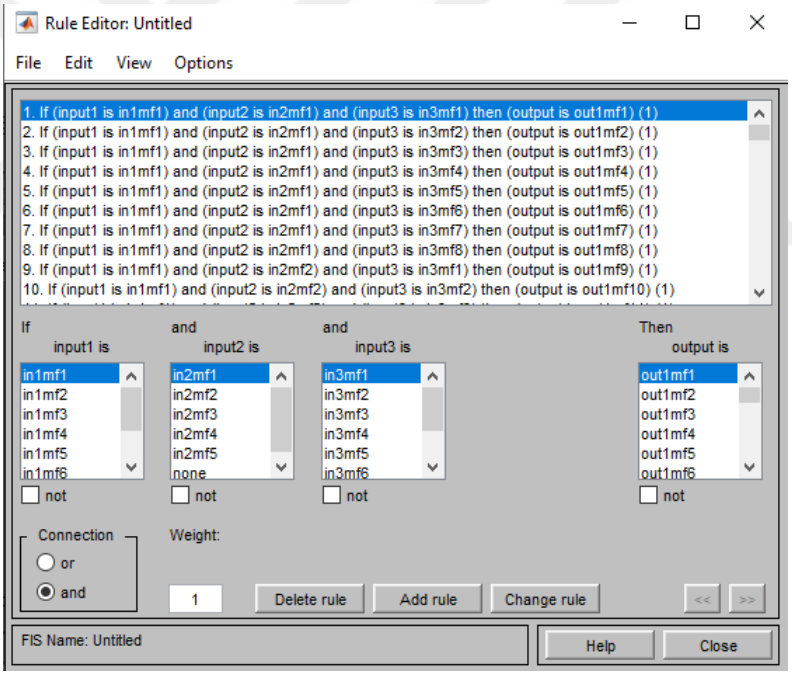


Figure 4.8 Rule editor of designer

After modeling is finished, results can be seen on the initial ANFIS interface by graphing against test and training data sets with model outputs. A surface graph can also be used to obtain the distribution of the final model with two inputs and one output on a 3D graph, as seen in Figure 4.9.

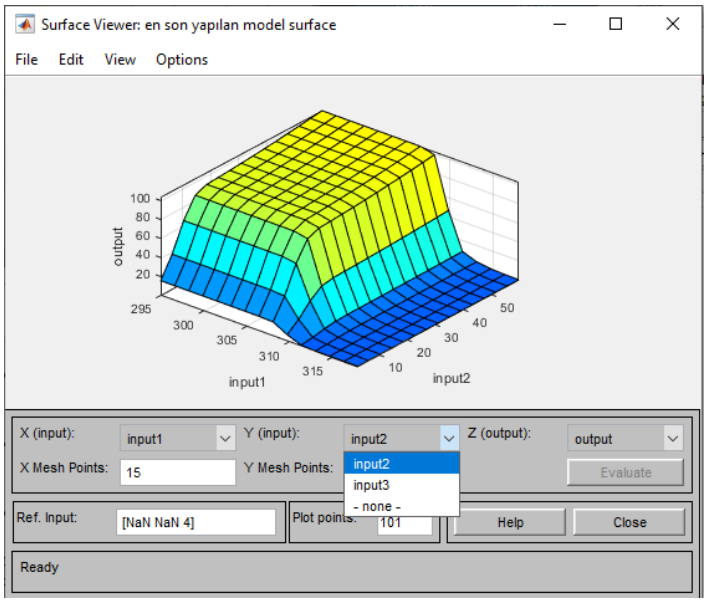


Figure 4.9 Surface graph of final model

4.2.1 Estimation of adsorption percentage via ANFIS

In this study, adsorption data obtained at different pH, carbon dosage, and temperature values were evaluated using MATLAB ANFIS. Train data given in Table 4.1 were used to train the program, and test data were used to evaluate the estimations. Four different MFs and numerous MF numbers can be examined for this estimation, and, depending on the type and number of MF, different predictions with a variety of errors can be obtained. There is limited information in the literature on how to choose the correct MF and MF numbers for adsorption data. Hence, the optimum MF and MF number should be obtained. As a result, experimental design methodology was used to optimize MF and MF number.

Adsorption data have three inputs (pH, carbon dosage, and temperature). Accordingly, for each parameter MF and the number of MF have to be chosen. In order to determine the optimum number of MF, the Box-Behnken experimental design method was applied.

There are numerous possibilities for the number of MF. In order to limit these possibilities, firstly, for each parameter, the same number of MF was chosen for each input as 5-5-5, 10-10-10 and 15-15-15, which also represents the center points in the Box-Behnken experimental design matrix. For each case, the error values of the training and testing data sets were recorded. Then, by examining the error values,

center points and MF types were chosen for the further experimental design evaluations.

4.2.2 Determination of optimum MF and the number of MF via Box-Behnken experimental design method

A typical three-parameter Box-Behnken experimental design matrix can be seen in Table 4.2. The levels corresponded the codded values are shown in Table 4.3. In order to determine the optimum number of MF for each input, all combinations given in Table 4.2 were examined via ANFIS, and the error (RMSE) values of training and testing data sets were recorded. After obtaining the error values from ANFIS, these values were transferred into the Design Expert software, and the relationship between the MF numbers for each input and the error values for training and testing data were evaluated using ANOVA. Accordingly, the most appropriate MF and the optimum combination for the number of MF were determined.

Table 4.2 Box-Behnken design matrix

No	Number of MF for Input 1	Number of MF for Input 2	Number of MF for Input 3
1	-1	0	-1
2	-1	0	1
3	-1	1	0
4	0	-1	1
5	0	1	-1
6	0	0	0
7	1	0	-1
8	0	-1	-1
9	1	1	0
10	0	0	0
11	1	0	1
12	1	-1	0
13	-1	-1	0
14	0	0	0
15	0	1	1

Table 4.3 Coded values for Box-Behnken experimental design

Parameters	Levels		
	-1	**0**	**1**
MF number for Input 1	2	5	8
MF number for Input 2	2	5	8
MF number for Input 3	2	5	8

5
RESULTS AND DISCUSSION

5.1 Determination of MF Types

By using the same number of MF for three inputs and the most used ten MF types, train adsorption data and test data were examined via ANFIS and then, error values were obtained for each data. The mean absolute error values obtained for each MF and MF numbers can be seen in Table 5.1.

Table 5.1 RMSE for train and test data obtained at different MF numbers

MF Type	Number of MF (Input 1- Input 2- Input 3)					
	5-5-5		10-10-10		15-15-15	
	RMSE		RMSE		RMSE	
	Train	Test	Train	Test	Train	Test
Trimf linear	0.29836	74.30540	0.00039	33.63160	0.00048	48.66430
Trimf cons	0.83764	3.68980	0.03964	39.07820	0.00011	48.79680
Trapmf linear	0.32057	33.49040	0.00036	48.27560	0.00031	77.81740
Trapmf cons	3.07220	35.43130	0.03544	47.45490	0.00011	48.72450
Gbellmf linear	0.00579	252.09110	0.00212	68.70180	0.00058	48.05020
Gbellmf cons	0.72650	4.08420	0.00444	55.54230	0.00013	47.31220
Gaussmf linear	0.00202	199.85640	0.00063	51.62050	0.00058	48.05020
Gaussmf cons	0.24849	3.46980	0.03465	41.47770	0.00012	48.30940
Gauss2mf linear	0.06116	2471.16000	0.00078	48.57450	0.00083	55.70180
Gauss2mf cons	0.51320	5.00120	0.02616	43.20830	0.00012	48.52450

Table 5.1 shows that as the number of MF increases error value for train data reduces but test error increases. It can be concluded that five MF number could be the center point for the for the further experimental design investigations. As can be seen in Table 5.1, constant type of MF showed the smaller error values, compared to linear functions. Both mean error values of train and test data are more appropriate than the values obtained with the linear functions, except trapmf constant function. Main reason to use center point is to get a general idea for MF type and the number. As a conclusion, the number of MF should be around 5 and MF should be the constant type. For the next step, chosen MF types were used in ANFIS. ANFIS starts by fuzzification of inputs, makes calculations then estimation values achieved by de-fuzzification done by constant or linear model. De-fuzzification depends on the related data set, in this study constant output function achieved better results.

5.2 Determination of Optimum MF Numbers via Experimental Design Approach

Mean error values of training and testing data sets obtained with different MF types and numbers (Table 4.2 and 4.3) can be seen in Table 5.2.

Table 5.2 RMSE obtained from ANFIS via BBD approach

No	MF number (Input 1-Input 2-Input 3)	Trimf Cons RMSE		Trapmf Cons RMSE	
		Train	Test	Train	Test
1	2-5-2	2.32840	4.33390	1.90160	5.34370
2	2-5-8	2.13970	5.04440	3.23480	35.55460
3	2-8-5	1.41600	13.91940	1.07840	32.21050
4	5-2-8	5.02100	7.46850	6.15900	7.50730
5	5-8-2	1.25550	3.50290	1.76390	5.22090
6	5-5-5	0.83764	3.68980	3.07220	35.43130
7	8-5-2	1.50130	3.80540	1.79950	4.94660
8	5-2-2	5.05550	7.68960	6.20050	7.87840
9	8-8-5	0.56624	12.47390	0.33176	35.16280
10	5-5-5	0.83764	3.68980	3.07220	35.43130
11	8-5-8	0.83770	3.69050	3.07220	35.43130
12	8-2-5	5.02110	7.46850	6.15900	7.50730
13	2-2-5	5.38900	7.75000	6.25220	7.75300
14	5-5-5	0.83764	3.68980	3.07220	35.43130
15	5-8-8	0.56610	12.49220	0.33176	35.16270

No	MF number (Input 1-Input 2-Input 3)	Gbellmf Cons RMSE		Gaussmf Cons RMSE		Gauss2mf Cons RMSE	
		Train	Test	Train	Test	Train	Test
1	2-5-2	1.87960	4.92150	2.01170	4.58370	1.92860	4.54590
2	2-5-8	1.62610	6.69360	1.18440	11.53130	0.86934	5.93300
3	2-8-5	0.52898	162.79170	0.90326	113.20200	0.75513	24.31910
4	5-2-8	15.50460	9.33210	5.49800	7.56680	5.05790	7.88700
5	5-8-2	0.49922	14.39470	1.13980	2.92160	0.88656	3.45880
6	5-5-5	0.72650	4.08420	0.24849	3.46980	0.51320	5.00120
7	8-5-2	1.40100	3.70370	1.43540	3.66470	1.34930	3.59790
8	5-2-2	15.84420	9.88460	5.50720	7.75650	5.10010	8.13040
9	8-8-5	0.01545	32.96380	0.41440	15.21230	0.03989	17.11460
10	5-5-5	0.72650	4.08420	0.24849	3.46980	0.51320	5.00120
11	8-5-8	0.72718	4.09620	0.24848	3.47130	0.51319	5.00280
12	8-2-5	15.50460	9.33210	5.47850	7.55610	5.05060	7.88850
13	2-2-5	15.82900	9.42930	5.53800	7.84480	5.08310	8.16520
14	5-5-5	0.72650	4.08420	0.24849	3.46980	0.51320	5.00120
15	5-8-8	0.01561	33.04860	0.41526	14.92690	0.05386	16.33750

For each selected MF type, Box-Behnken design approach was applied and obtained results transferred to Design Expert software. It was aimed to develop a regression model between the mean error values and MF number. Developed quadratic models

for each MF types were given in Table 5.3. R1 and R2 show the mean error obtained with train data and test data, respectively, and the subscripts indicate the MF type. Developed models give the relationship between the MF number of each input (MF1, MF2, MF3, for input 1, input2 and input3, respectively) and error values, which would be obtained via ANFIS. As can be seen from Table 5.3, depending on MF type, the MF numbers of each input has different coefficient that means different effect on error. It can be concluded that, the optimum numbers of MF for each input should be determined depending on the used MF type.

Table 5.3 Developed regression models

MF Type		Equation
Trimfcons	**Train**	$R1_{Trimfcons}$=+11.5272 -0.5553xMF1 -2.5001xMF2 -0.3202xMF3 -0.0134xMF1 x MF2 -0.0132xMF1 x MF3 -0.0182xMF2 x MF3 +0.0549x$MF1^2$ +0.1963x$MF2^2$ +0.0411x$MF3^2$
	Test	$R2_{Trimfcons}$= +21.3268-1.7351xMF1-6.3298xMF2+0.2698xMF3-0.0323xMF1 x MF2+0.2558xMF2 x MF3+0.1746x$MF1^2$+0.5712x$MF2^2$-0.1159x$MF3^2$
Trapmfcons	**Train**	$R1_{Trapmf\ cons}$=+6.5275 +0.4579xMF1 -1.4326xMF2 +0.4775xMF3 -0.0182xMF1 x MF2 -0.0017xMF1 x MF3 -0.0386xMF2 x MF3 -0.0405x$MF1^2$ +0.0831x$MF2^2$ -0.0229x$MF3^2$
	Test	$R2_{Trapmfcons}$=-47.7858 +4.2734xMF1 +10.3085xMF2 +11.6398xMF3 +0.0888xMF1 x MF2 +0.0076xMF1 x MF3 +0.8420xMF2 x MF3 -0.4665x$MF1^2$ -1.1750x$MF2^2$ -1.2127x$MF3^2$
Gbellmfcons	**Train**	$R1_{Gbellmfcons}$=+34.9293 -0.3885xMF1 -10.1883xMF2 -0.3714xMF3 -0.0053xMF1 x MF2 -0.0117xMF1 x MF3 -0.0040xMF2 x MF3 +0.0381x$MF1^2$ +0.7667x$MF2^2$ +0.0377x$MF3^2$
	Test	$R2_{Gbellmfcons}$=-17.7657 -8.5176xMF1 -7.5178xMF2 +20.9526xMF3 -23.6036xMF1 x MF2 +2.0963x$MF1^2$ +3.4087x$MF2^2$ -2.0108x$MF3^2$
Gaussmfcons	**Train**	$R1_{Gaussmfcons}$=+13.4979 -0.4846xMF1 -3.2806xMF2 -0.5363xMF3 -0.0119xMF1 x MF2 -0.0010xMF1 x MF3 -0.0199xMF2 x MF3 +0.0508x$MF1^2$ +0.2641x$MF2^2$ +0.0571x$MF3^2$
	Test	$R2_{Gaussmfcons}$=-14.9025-7.5683xMF1 -1.0409xMF2 +14.8392xMF3 -2.7139xMF1 x MF2 +1.6669x$MF1^2$ +1.9424x$MF2^2$ -1.4066x$MF3^2$
Gauss2mfcons	**Train**	$R1_{Gauss2mfcons}$ =+11.5670 -0.3449xMF1 -2.6955xMF2 -0.4224xMF3 -0.0190xMF1 x MF2 +0.0062xMF1 x MF3 -0.0220xMF2 x MF3 +0.0339x$MF1^2$ +0.2127x$MF2^2$ +0.0386x$MF3^2$
	Test	$R2_{Gauss2mfcons}$ =+20.1466 -2.3095xMF1 -7.1755xMF2 +1.9590xMF3 -0.1924xMF1 x MF2 +0.3645xMF2 x MF3 +0.2882x$MF1^2$ +0.7530x$MF2^2$ -0.3139x$MF3^2$

In order to evaluate the developed models for each MF type, ANOVA analysis was used via Design Expert software. The method of least squares is typically used to estimate the regression coefficients. P-value is used to identify process parameters, which has statistically significant effect on each response. P-values less than 0.05 indicate that parameters are significant in 95% confidence interval. Table 5.4 shows ANOVA results for developed models. As can be seen from Table 5.4, developed models are mostly significant (with high R^2 values) except gbellmf cons and gauss2mf cons MF types.

Table 5.4 ANOVA results

MF Type		Sum of Squares	df	Mean Square	F-value	p-value	R^2	Significancy
trimf cons	Train	48.88	9	5.43	247.70	< 0.0001	0.9978	Significant
	Test	163.57	8	20.45	5.61	0.0249	0.882	Significant
trapmf cons	Train	60.36	9	6.71	15.31	0.0039	0.9650	Significant
	Test	2812.98	9	312.55	7.29	0.0207	0.9292	Significant
gbellmf cons	Train	651.78	9	72.42	9528.88	< 0.0001	0.9999	Significant
	Test	17993.87	7	2570.55	3.69	0.0533	0.7867	Not Significant
gaussmf cons	Train	69.03	9	7.6	116.63	< 0.0001	0.9953	Significant
	Test	8195.05	7	1170.72	3.18	0.0749	0.7608	Significant
gauss2mf cons	Train	58.40	9	6.49	237.83	< 0.0001	0.9977	Significant
	Test	433.14	9	48.13	3.24	0.1045	0.8535	Not Significant

Optimization module in Design Expert Software was used, in order to obtain optimum MF numbers for each input (MF1, MF2 and MF3) which gave the minimum errors (R1 and R2) for train and test data. Accordingly, obtained optimum MF numbers (after rounding to the nearest integer) for each MF type were given in Table 5.5. As can be seen that different optimum MF numbers were obtained for different MF types. ANFIS prediction was also examined by using obtained optimum

MF numbers, then, errors between the predicted data by ANFIS and experimental data were determined. It can be said that these errors are in consistent with the values obtained from developed models except gbellmf cons MF type (Table 5.5).

Table 5.5 ANFIS results and ANOVA prediction comparison of selected MFs

No	MF Type	Optimum MF number			ANFIS Prediction		Regression model prediction	
					RMSE		RMSE	
		MF 1	MF2	MF 3	TRAINING	TEST	TRAINING	TEST
1	trimf cons	6	6	3	0.64882	1.9084	0.6	2.998
2	trapmf cons	8	8	2	1.7639	5.2209	0.511	1.527
3	gbellmf cons	7	7	8	0.037171	37.9536	-1.369	2.082
4	gaussmf cons	7	6	7	0.1722	3.3029	-0.358	-1.551
5	gauss2mf cons	6	6	3	0.64344	3.4153	0.314	3.405

After this point, the most fitting model should be picked by the user. Regression model prediction for trimf cons matches ANFIS very well, as expected. Accordingly, adsorption test data predicted by ANFIS with optimum MF numbers of 6, 6 and 3 for input 1,2 and 3 (temperature, carbon dosage and pH) were compared with experimental adsorption data as seen in Table 5.6. Adsorption percentage predicted by ANFIS with optimum conditions are in good agreement with the experimental values.

Error values in table 5.6 was calculated by following equation;

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \tag{5.1}$$

Table 5.6 ANFIS prediction results obtained optimum MF and MF numbers

Test data	Temp.° C	Carbon dosage (mg)	pH	Adsorption (%)		Error (%)
				Experimental	ANFIS predicted	
1	295	7.8	2	63.95	61.3	4.3230
2	295	17.5	2	98.98	99.8	0.8216
3	295	10.3	3	67.73	71.0	4.6056
4	295	29.7	3	99.89	101.0	1.0990
5	295	8.7	4	53.78	53.6	0.3358
6	295	30.8	4	99.64	97.7	1.9857
7	295	11.2	6	32.43	32.3	0.4025
8	305	10.5	2	67.61	69.4	2.5793
9	318	10.4	2	78.22	75.6	3.4656
				RMSE		1.9195

6
CONCLUSION

ANFIS is a great tool to make estimations. In order to achieve results with acceptable error values, determining MF and the number of MF are crucial steps. The general approach to apply ANFIS achieved by modelling an experimental adsorption data set. The goal of this study was to identify a general approach by optimizing the number of MF for ANFIS prediction.

Following conclusions can be drawn from this study:

- ANFIS prediction was made over a very wide error range depending on MF type and MF number.

- For the adsorption data set, constant types of MFs showed acceptable predictions.

- Comparing the number of MF, the level of 5 was resulted in lower errors than that of level 15.

- Optimum number of MF for ANFIS prediction was determined by using experimental design approach.

- For different MF type, optimum number of MF for each input could be different.

- Adsorption test data were predicted well via ANFIS with trimf and optimum MF numbers of 6, 6 and 3 for input 1,2 and 3 (temperature, carbon dosage and pH).

- Adsorption percentage predicted by ANFIS with optimum conditions were in good agreement with the experimental values.

REFERENCES

Aslan, N., & Cebeci, Y. (2007). Application of box–behnken design and response surface methodology for modeling of some Turkish coals. Fuel, 86(1-2), 90–97. https://doi.org/10.1016/j.fuel.2006.06.010

Duranoglu D., Trochimczuk A. W., Beker U.(2010). A comparison study of peach stone and acrylonitrile-divinylbenzene copolymer based activated carbons as chromium(VI) sorbents. CHEMICAL ENGINEERING JOURNAL, vol.165, pp.56-63, https://doi.org/10.1016/j.cej.2010.08.054

Dutta, P., & Kumar, A. (2018). Application of an ANFIS model to Optimize the Liquid Flow Rate of a Process Control System. Chemical Engineering Transactions, 71, 991-996. https://doi.org/10.3303/CET1871166

Fan, B., Lin, C., Wang, F., Liu, S., Liu, L., & Xu, S. (2018). An adaptive neuro-fuzzy inference system (ANFIS) based model for the temperature prediction of lithium-ion power batteries. SAE International Journal of Passenger Cars - Electronic and Electrical Systems, 12(1). https://doi.org/10.4271/07-12-01-0001

Goodarzi, M., & Freitas, M. P. (2010). Mia–QSAR coupled to principal component analysis-adaptive neuro-fuzzy inference systems (PCA–ANFIS) for the modeling of the anti-HIV reverse transcriptase activities of tibo derivatives. European Journal of Medicinal Chemistry, 45(4), 1352–1358. https://doi.org/10.1016/j.ejmech.2009.12.028

Hong, E., Yeneneh, A. M., Sen, T. K., Ang, H. M., & Kayaalp, A. (2018). ANFIS based modelling of dewatering performance and polymer dose optimization in a wastewater treatment plant. Journal of Environmental Chemical Engineering, 6(2), 1957–1968. https://doi.org/10.1016/j.jece.2018.02.041

Kanwal, S., Khan, B., Ali, S. M., & Mehmood, C. A. (2018). Gaussian process regression based inertia emulation and reserve estimation for grid interfaced photovoltaic system. Renewable Energy, 126, 865–875. https://doi.org/10.1016/j.renene.2018.04.012

Khoshvaght, H., Delnavaz, M., & Leili, M. (2021). Optimization of acetaminophen removal from high load synthetic pharmaceutical wastewater by experimental and ANOVA analysis. Journal of Water Process Engineering, 42, 102107. https://doi.org/10.1016/j.jwpe.2021.102107

Omar, A., Aous, A. & Balasem, S. (2015). Comparison between the Effects of Different Types of Membership Functions on Fuzzy Logic Controller Performance. International Journal of Emerging Engineering Research and Technology Volume 3, Issue 3, March 2015, PP 76-83

Onu, C. E., Nwabanne, J. T., Ohale, P. E., & Asadu, C. O. (2021). Comparative analysis of RSM, Ann and ANFIS and the mechanistic modeling in eriochrome black-T dye adsorption using modified Clay. South African Journal of Chemical Engineering, 36, 24–42. https://doi.org/10.1016/j.sajce.2020.12.003

Qin, J., Du, S., Ye, J., & Yong, R. (2022). SVNN-ANFIS approach for stability evaluation of open-pit mine slopes. Expert Systems with Applications, 198, 116816. https://doi.org/10.1016/j.eswa.2022.116816

Rust, A., Marini, F., Allsopp, M., Williams, P. J., & Manley, M. (2021). Application of ANOVA-simultaneous component analysis to quantify and characterise effects of age, temperature, syrup adulteration and irradiation on near-infrared (NIR) spectral data of Honey. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 253, 119546. https://doi.org/10.1016/j.saa.2021.119546

Sarkar, S., Chowdhury, R., Das, R., Chakraborty, S., Choi, H., & Bhattacharjee, C. (2014). Application of ANFIS model to optimise the photocatalytic degradation of chlorhexidine digluconate. RSC Advances, 4(40), 21141. https://doi.org/10.1039/c4ra00389f

Soltany Mahboob, A., & Zahiri, S. H. (2019). Variable length IPO and its application in concurrent design and train of ANFIS systems. Applied Intelligence, 49(6), 2233–2255. https://doi.org/10.1007/s10489-018-1366-0

Sreekumar, S., Kallingal, A., & Mundakkal Lakshmanan, V. (2020). Adaptive neuro-fuzzy approach to sodium chlorate cell modeling to predict cell ph for energy-efficient chlorate production. Chemical Engineering Communications, 208(2), 256–270. https://doi.org/10.1080/00986445.2019.1708740

Talpur, N., Salleh, M. N., & Hussain, K. (2017). An investigation of membership functions on performance of ANFIS for solving classification problems. IOP Conference Series: Materials Science and Engineering, 226, 012103. https://doi.org/10.1088/1757-899x/226/1/012103

Tanhaei, B., Esfandyari, M., Ayati, A., & Sillanpää, M. (2016). Neuro-fuzzy modeling to adsorptive performance of magnetic chitosan nanocomposite. Journal of Nanostructure in Chemistry, 7(1), 29–36. https://doi.org/10.1007/s40097-016-0211-4

Teimouri, R., & Sohrabpoor, H. (2013). Application of adaptive neuro-fuzzy inference system and cuckoo optimization algorithm for analyzing Electro Chemical Machining Process. Frontiers of Mechanical Engineering, 8(4), 429–442. https://doi.org/10.1007/s11465-013-0277-3

Wang, H., Luan, W., Sun, L., Zeng, Z., Xue, W., & Bai, Y. (2022). Study on polyvinyl butyral purification process based on box-behnken design and Artificial Neural Network. Chemical Engineering Research and Design, 184, 291–302. https://doi.org/10.1016/j.cherd.2022.05.050

Wikimedia Foundation. (2022, December 24). Fuzzy logic. Wikipedia. Retrieved December 28, 2022, from https://en.wikipedia.org/wiki/Fuzzy_logic

Wong, Y. J., Arumugasamy, S. K., Chung, C. H., Selvarajoo, A., & Sethu, V. (2020). Comparative study of artificial neural network (ANN), adaptive neuro-Fuzzy Inference System (ANFIS) and multiple linear regression (MLR) for modeling of Cu (II) adsorption from aqueous solution using biochar derived from rambutan (Nephelium lappaceum) peel. Environmental Monitoring and Assessment, 192(7). https://doi.org/10.1007/s10661-020-08268-4

Wu, D. (2012). Twelve considerations in choosing between Gaussian and trapezoidal membership functions in interval type-2 fuzzy logic controllers. 2012 IEEE International Conference on Fuzzy Systems. https://doi.org/10.1109/fuzz-ieee.2012.6251210

Zaghloul, M. S., Hamza, R. A., Iorhemen, O. T., & Tay, J. H. (2020). Comparison of adaptive neuro-fuzzy inference systems (ANFIS) and support vector regression (SVR) for data-driven modelling of aerobic granular sludge reactors. Journal of Environmental Chemical Engineering, 8(3), 103742. https://doi.org/10.1016/j.jece.2020.103742

[1] Clusteringtype. Option set for genfis function - MATLAB. (n.d.). Retrieved January 7, 2023, from https://www.mathworks.com/help/fuzzy/genfisoptions.html

PUBLICATIONS FROM THE THESIS

Conference Papers

1. ALTIN, E.S., KÜÇÜK, İ., DURANOĞLU, D. (2022). Application of Adaptive Neuro-Fuzzy Inference System (ANFIS) on an Adsorption Process. 4th International Eurasian Conference on Science, Engineering and Technology, 179, ISBN: 978-605-72134-1-9