Electrical Engineering
https://doi.org/10.1007/s00202-020-00951-6

ORIGINAL PAPER

Fuzzy system applied to a hydraulic turbine efficiency curve fitting

Ramon Abritta[1] · Frederico F. Panoeiro[1] · Eduardo P. de Aguiar[1] · Leonardo de M. Honório[1] · André L. M. Marcato[1] · Ivo C. da Silva Junior[1]

Received: 26 November 2019 / Accepted: 12 February 2020

Abstract
This paper reports the results of different fuzzy-based approaches applied to the fitting of a hydraulic turbine efficiency curve. This kind of curve is granted by the turbine manufacturer as a three-dimensional dataset that needs to be properly fitted in order to provide the turbine efficiency for any values of net head and water discharge in the relevant space and, therefore, guarantee an as realistic as possible representation of the hydroelectric power plant's machines. The clustering algorithm Fuzzy C-Means and the ANFIS and Extreme Learning ANFIS architectures were widely tested and compared to the conventional polynomial adjustment. Since the studied curve is nonlinear, researches involving any kind of nonlinear curve fitting can benefit from this work's information.

Keywords Curve fitting · Fuzzy system · Clustering · Extreme learning machine · ANFIS

1 Introduction

The Unit Commitment Problem comes down to the optimal allocation of generating units in order to minimize costs [2,28], minimize resources usage [9,15], maximize efficiency [4,13] or maximize profit [10] during the planning horizon. In a hydroelectric power plant (HPP) optimization, different aspects of the system are mathematically modeled, such as the penstock losses coefficient, the generators' efficiency curve, the turbines' efficiency curve (TEC), among others, depending on the adopted HPP model. Since most of Brazilian's power generation comes from HPPs, it is of great importance that their demand attendance is optimized [11,14,40,42] in order to guarantee minimum losses, resources usage, among other goals, depending on the model's objective function.

When addressing the optimal dispatch problem (ODP), it is desired that the HPP parameters and characteristics, such as the TEC, be represented as close as possible to reality [17–19], so that the model's provided solution is more coherent, and more accurate results can be secured. Within this matter, regarding the TEC, it is necessary to determine a function/black box that returns the turbine efficiency (n_t) as output for any values of net head (N_h) and water discharge (q) inputs. Such curves are originally provided as a set data of points; therefore, obtaining a proper fitted function/black box is not trivial. Following such a line of thinking, this paper presents a novel approach to the fitting of a real TEC, which aims at granting a smooth surface with small errors regarding the dataset. Data of the machines from the HPP Luís Eduardo Magalhães, located in Lajeado town, Tocantins State, Brazil, were utilized to perform the simulations.

Polynomials are commonly applied to the aforementioned problem, i.e., the fitting of the TECs. As an advantage, their derivatives are easily determined, which may be helpful during the optimization implementation. Little exploration of different methodologies to the approached subject can be found in the literature. In fact, the authors of this paper could not find any work that aimed specifically at studying distinct ways of obtaining a function/black box to model the tur-

✉ Ramon Abritta
ramon.abritta@engenharia.ufjf.br

Frederico F. Panoeiro
frederico.panoeiro@engenharia.ufjf.br

Eduardo P. de Aguiar
eduardo.aguiar@engenharia.ufjf.br

Leonardo de M. Honório
leonardo.honorio@ufjf.edu.br

André L. M. Marcato
andre.marcato@ufjf.edu.br

Ivo C. da Silva Junior
ivo.junior@ufjf.edu.br

1 Graduate Program in Electrical Engineering, Federal University of Juiz de Fora, Juiz de Fora, MG, Brazil

Published online: 28 February 2020

bines' efficiency. In this work, the Fuzzy C-means (FCM) algorithm, the Adaptive-Network-based Fuzzy Inference System (ANFIS) and Extreme Learning Adaptive-Network-based Fuzzy Inference System (ELANFIS) architectures were explored in the matter. The following paragraphs concisely describe the aforementioned techniques.

Clustering techniques are used in many research areas, e.g., image segmentation [35,37], pattern recognition [34], classification [12,45] and data mining [43]. These algorithms group data that have similar characteristics. Many of these techniques are based on the fuzzy system (FS), such as the FCM [3] and its variants: Possibilistic C-Means [26], Fuzzy C-Numbers [46], Fuzzy Possibilistic C-Means [30], Entropy-Regularized FCM [27], Partition Simplification FCM [23], Interval Type 2 Fuzzy C-Means [36], Possibilistic Fuzzy C-Means [31], Single-Pass Fuzzy C-Means [20], Interval Type 2 Possibilistic C-Means [47], Random Sampling Plus Extension Fuzzy C-Means [16], Geometric Progressive Fuzzy C-Means [32], Minimum Sample Estimate Random Fuzzy C-Means [32] and Interval Type 2 Fuzzy Possibilistic C-Means [38].

The ANFIS architecture [25] is a hybrid system that combines Takagi–Sugeno–Kang fuzzy model's knowledge representation, uncertainties and impreciseness tolerance, and interpretability with artificial neural network's (ANN) adaptability and learning capability aiming at better performance. Several research areas utilize ANFIS, e.g., restoration of images [8], thermal power plant control [5] and transmission lines fault location [1].

Extreme learning machine (ELM) [22] is a derivation of single-hidden-layer feed-forward ANN that greatly increases learning speed by using a least-square method to approximate the weights to their optimal values at each activation function, thus avoiding the hidden layer parameters tuning [7]. Applications include soft sensors [48], classification [41] and emotion recognition [21]. Misclassifications may occur due to the random generation of the input hidden node parameters and the fact that these remain unchanged during training [6].

Extreme-ANFIS (ELANFIS) [33] is a combination of ELM and ANFIS that incorporates the FS to the architecture, thus adding knowledge representation to it and consequently avoiding ELM's intrinsic randomness. Moreover, it eliminates ANFIS's hybrid learning algorithm, hence reducing computational complexity. Some applications are regression problems [24], control [39], classification [44] and chaotic time series prediction [29]. Both ANFIS and ELANFIS parameters to be set are the number of membership functions (N_{F}) and the number of training epochs (N_{E}).

Regarding this work's organization, Sect. 2 describes the problem, Sect. 3 presents distinct fitting approaches, Sect. 4 compares these approaches and Sect. 5 exhibits a brief conclusion and possible further works.

2 Problem description

Turbine efficiency curves (TECs) are usually provided as a graph in which the axes represent the HPP N_{h} in meters and q in cubic meters per second, whereas n_{t} levels are given as lines over the graph. Therefore, the first step in the fitting process is to extract the data from the mentioned graph. Figure 1 shows the extracted samples from HPP Luís Eduardo Magalhães' TEC. A few observations can be done concerning TECs in general: (i) the surface is nonlinear; (ii) although some regions have high concentration of data, others present considerable voids, which can be observed in Fig. 1; and (iii) the data points are not chaotically spread over the space;

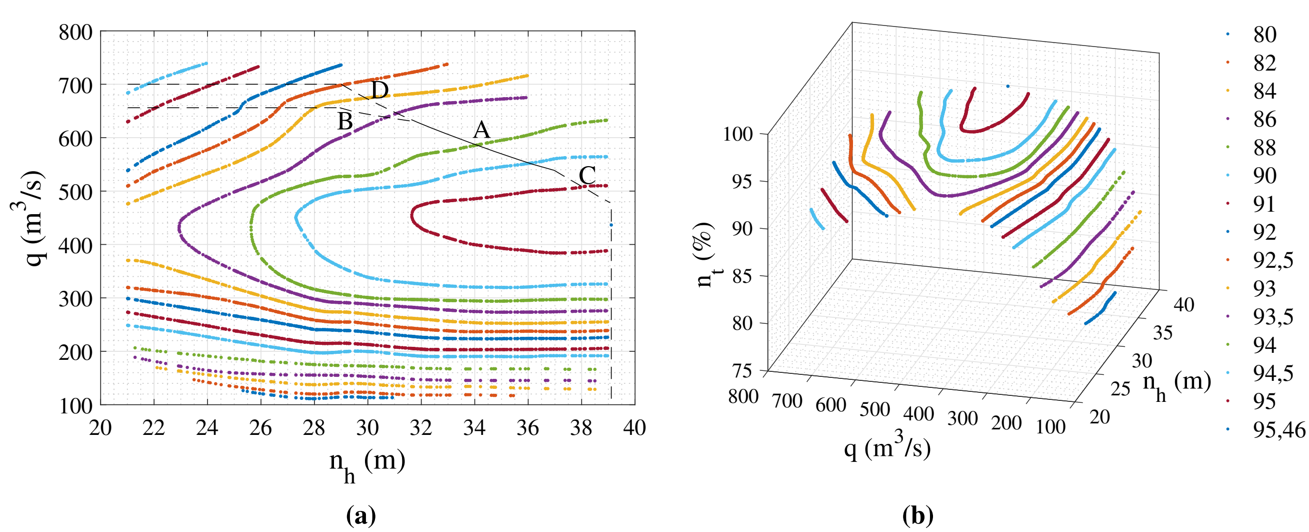


Fig. 1 Samples extracted from HPP Luís Eduardo Magalhães' TEC plotted in **a** two dimensions (including CL) and **b** three dimensions

in fact, they are disposed as relatively well-behaved lines. The curves A, B, C and D in Fig. 1a are the cavitation limits (CL), B being for during guarantee period and D for after guarantee. All TECs have CL, which determine the so-called forbidden zone. Since operating in this zone severely compromises the machines' life span, such operations are always avoided. Therefore, surface distortions beyond the CL have little to no relevance.

Regarding the given approach to the problem, the extracted data were fitted by several methods. Mean (e_{mn}) and maximum (e_{mx}) percentage relative errors were compared among them so that evaluations could be done. As a crucial step, each generated surface (GS) was analyzed, since undesirable distortions may occur, which will soon become clear. It is up to the specialist to decide what level of distortion makes a GS unviable.

As mentioned in Sect. 1, the HPP's equipment and parameters should be represented as closely to reality as possible since, if such a measure is taken, more accurate results are likely to be obtained by the ODP optimization model. This way, better support, and more precise reports can be offered to the plant operators. Although these affirmations relate to the operation of HPPs, any model that simulates real-life events/phenomena can benefit from a more precise representation. Within this context, this paper's results have shown that applying clustering to the TEC's data caused the fitting process to better absorb the nonlinearities of the curve. Furthermore, compared to polynomial fittings, the ones performed by ANFIS and ELANFIS provided less significant errors regarding the dataset. In conclusion, the proposed approach has granted a more accurate representation of the TEC. Details that led to this statement will be exposed and explained in the next sections.

Two major advantages are worth mentioning regarding the fitting process carried out by the proposed approach: (i) unlike polynomials fittings, which are given by strict mathematical rules and, therefore, always provide the same function for a given dataset, ANFIS, and ELANFIS fittings are based on procedures that simulate human intelligence. That means that the ANN-based approach benefits from adaptability, hence allowing the execution of multiple simulations, each providing a different network, i.e., attempts of obtaining better nets may be performed; and (ii) theoretically, a clustering algorithm such as the FCM is capable of properly grouping the data, which removes the necessity of developing a specific method for dividing data points between two or more sets.

The contributions of this paper are:

- An unexplored approach to TECs' dataset clustering is proposed, which utilizes the FS to accomplish the given task;
- Computational intelligence techniques (ANFIS and ELANFIS) are used to fit the TEC's dataset, which are tools not yet applied to this specific problem.

The main conclusions are:

- FS-based clustering has shown to be promising concerning the TEC fit problem;
- ANFIS and ELANFIS provided smaller errors than the regular polynomial fitting.

3 Fitting methods

It is common that papers reporting HPPs optimization results do not discourse specifically about the treatment given to the studied HPP's TEC. Actually, this subject is often ignored, probably due to the tradition of utilizing polynomials to fit the TEC data. Given the increasing and successful applications of artificial intelligence techniques in distinct fields of science, the authors of this paper decided to apply the methods exposed in Sect. 1, i.e., ANFIS and ELANFIS combined to the FCM algorithm, to the TEC fitting problem. It is important to emphasize that applying the aforementioned techniques to the present problem is still an unexplored path. For this reason, this paper can potentially inspire other researchers to test similar approaches in similar problems and obtain promising results.

In the following subsections, different methodologies utilized to fit a real TEC are described. Both ANFIS and ELANFIS were tested considering and not considering clustered data. Furthermore, polynomial fittings were also evaluated.

3.1 Single fitting

This subsection describes the methods applied to the whole curve at once, i.e., regardless of the method, only one function/black box was created to provide n_{t} for any N_{h} and q sample. As an advantage, using a single fitting ensures that no distortions or discontinuities are created on the GS. However, it may be a difficult task to fit all data at once while providing small errors.

As a more simple method, polynomials were tested. It is important to highlight that such an approach, i.e., using a single polynomial to obtain the definitive GS, is not advisable when optimizing real problems since it is not capable of properly representing the nonlinearities of the dataset. In this paper, it was applied only for the sake of comparison.

Many are the applications of ANFIS in several research areas. In this paper, this technique was applied to the exposed fitting problem and its results were analyzed. The parameters

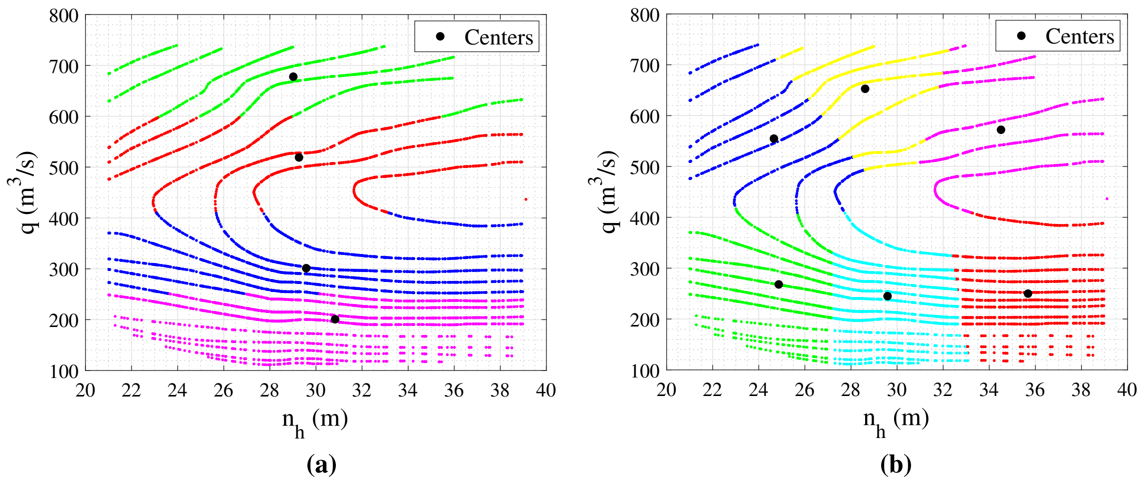


Fig. 2 TEC data clustered using **a** E_{D} with $N_{\mathrm{C}} = 4$ and **b** M_{D} with $N_{\mathrm{C}} = 6$

N_{F} and N_{E} were varied in order to verify their influence on GSs and errors.

The ELANFIS method, which is a combination of ELM and ANFIS, was also applied to the problem, and the same procedures described for the ANFIS application were performed.

3.2 Clustered data fitting

A common and efficient way to reduce errors in any curve fitting is to split the data into clusters and then fit each one separately. No matter the method used, the lesser the number of data samples, the easier the adjustment, as long as the data are properly grouped.

Many clustering methods can be found in the literature. Still concerning these, many are the mathematical principles that originate such methods. For instance, the ones exposed in Sect. 1 are based on the fuzzy concept. The FCM algorithm, which was the inspiration for all aforementioned algorithms, associates each data sample with one membership value (MV) from 0 to 1 for each cluster by minimizing a samples-centers distance-based objective function. The sum of these MVs is equal to 1 for every sample, which means that a particular sample does not exclusively belong to one cluster. In fact, it belongs to all of them with different relevancies. Each cluster has a center, for which is possible to calculate the distance to a sample that was not used in the clustering process and, by doing so, determine to what cluster this sample relates the most. In this work, this step is essential to obtain n_{t} for samples that are not part of the original dataset.

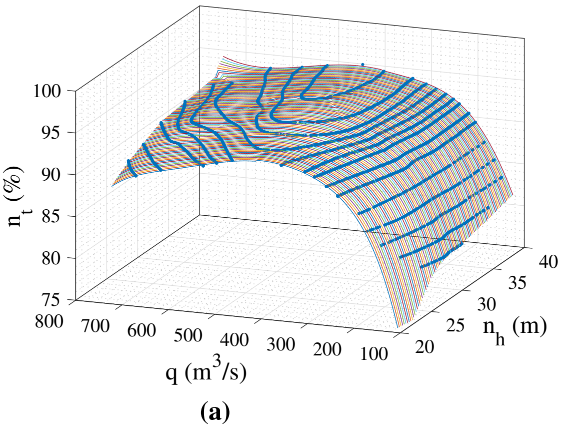
The FCM can be executed considering different kinds of distance in space. In this work, Euclidean (E_{D}) and Mahalanobis (M_{D}) distances were studied, and the number of clusters (N_{C}) was varied from 2 to 12. Figure 2 exemplifies some clustering possibilities considering both the aforementioned distance kinds. Once the dataset was clustered, tests with the different fitting methods used so far, i.e., polynomials, ANFIS and ELANFIS, were performed.

It is important to mention that the FCM algorithm utilizes a fuzziness parameter to determine how "nonhard" the variables are. In this work, this parameter was considered equal to 2.

A very important step of data clustering aimed at curve fitting is to overlap (OL) the clusters so that the GS presents improved smoothness and continuity. The fuzziness characteristic previously described is useful to establish OL between clusters. In this work, if the highest MV of a particular sample is greater than or equal to 0.7, it is considered to strongly belong to the corresponding cluster and, therefore, only with this cluster, it is associated. If the mentioned parameter is smaller than 0.7, it is associated not only with the cluster it relates the most but also with the cluster for which it has the second-highest MV. Lastly, if the third-highest MV of a sample is greater than 0.1, it is also associated with the cluster corresponding to this MV. In order to expose the importance of OL, Fig. 3 clarifies that not using OL data is pointless, since such an approach severely compromises the GS's integrity.

4 Experimental results

This section presents the analysis and error tables for all exposed approaches. In all tables presented, values branded with an apostrophe correspond to viable GSs. All simulations were performed in MATLAB ®2016a. The polynomial and ANFIS fittings were executed through the "fit" and "ANFIS" MATLAB commands. The ELANFIS codes may be found at https://github.com/PushpakJagtap/Extreme-ANFIS.



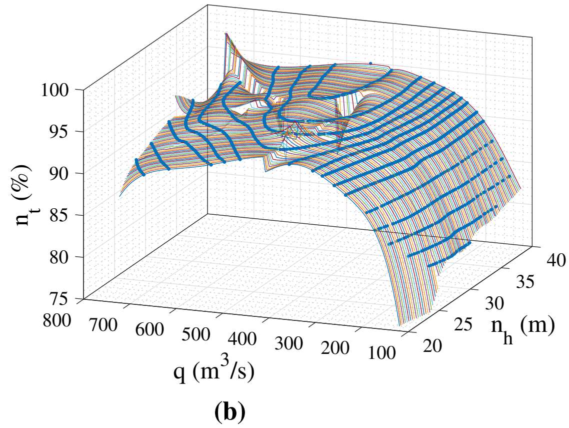


Fig. 3 Polynomials of degree 5 fitting 9 M_D FCM clusters **a** with and **b** without OL

Table 1 Single polynomial errors

Degree	2	3	4	5
e_{mn}	0.9284′	0.3593′	0.1869′	0.1424′
e_{mx}	6.7748′	3.0345′	1.7616′	1.3786′

For both nonclustered and clustered data, polynomials of degrees 2, 3, 4 and 5 were tested. For nonclustered data and for both ANFIS and ELANFIS, N_F was initially varied from 2 to 10 considering several N_E possibilities. For clustered data and for both ANFIS and ELANFIS, N_F was varied from 2 to 6 with a fixed value of 1000 N_E due to the heavy computational burden associated with the simulations. Table 1 lists the errors for the single polynomial fitting. As mentioned in Sect. 3, this method serves the sole purpose of providing comparisons.

Table 2 lists the errors for the single ANFIS and reveals that increasing N_F and N_E resulted in an overall tendency, with few exceptions, to reduce e_{mn}, whereas e_{mx} presented this same tendency, though with less consistency.

For N_F from 3 to 10, the GSs presented more expressive distortions beyond CL with the higher value of N_F. However, as mentioned in Sect. 2, these are not relevant. For N_F equal to 5, all GSs presented a minor, though compromising, distortion. For N_F equal to 6, GSs were viable for N_E equal to 100 and 200. However, they presented minor, though compromising, distortions for N_E equal to 500 and 1000. For N_F from 7 to 10, GSs presented graver relevant distortions with the higher value of N_F for all N_E, as exemplified in Fig. 4.

Other nets with higher N_E were trained for N_F from 2 to 6 in order to verify this parameter's influence in errors and in GSs with N_F equal to 5 and 6. Such an experiment did not affect the GS's integrity in N_F equal to 2–4 cases. For N_F equal to 5 and 6, the minor distortions remained with higher N_E. Table 3 lists the intrinsic errors and exposes the fact that increasing N_E resulted in a consistent, though small, or even null, reduction to their values.

As for the single ELANFIS approach, Table 4 lists the errors and reveals that increasing N_F very often resulted in lower e_{mn}, whereas e_{mx}, despite behaving inconsistently, presented a tendency to be reduced. Increasing N_E reduced e_{mn} in many cases. However, such an increase affected e_{mx}in a highly unpredictable way. The GSs provided by ELANFIS

Table 2 Single ANFIS errors

N_F	$N_E = 100$		$N_E = 200$		$N_E = 500$		$N_E = 1000$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
2	0.1682′	1.8563′	0.1467′	2.0038′	0.1447′	2.0325′	0.1447′	2.0324′
3	0.1416′	1.4162′	0.1205′	1.3049′	0.1062′	1.2632′	0.1053′	1.2652′
4	0.1464′	2.0095′	0.1377′	1.9567′	0.1328′	1.9273′	0.1326′	1.9232′
5	0.0865	1.3913	0.0844	1.3750	0.0833	1.3655	0.0833	1.3654
6	0.0943′	1.1985′	0.0869′	1.2096′	0.0573	1.0856	0.0567	1.0690
7	0.0504	0.8106	0.0405	0.6441	0.0393	0.6188	0.0389	0.6050
8	0.0533	1.0608	0.0517	1.0313	0.0508	1.0116	0.0507	1.0102
9	0.0410	0.9386	0.0317	1.1426	0.0263	0.9723	0.0296	0.7332
10	0.0306	0.8637	0.0247	0.9017	0.0245	0.8854	0.0244	0.8837

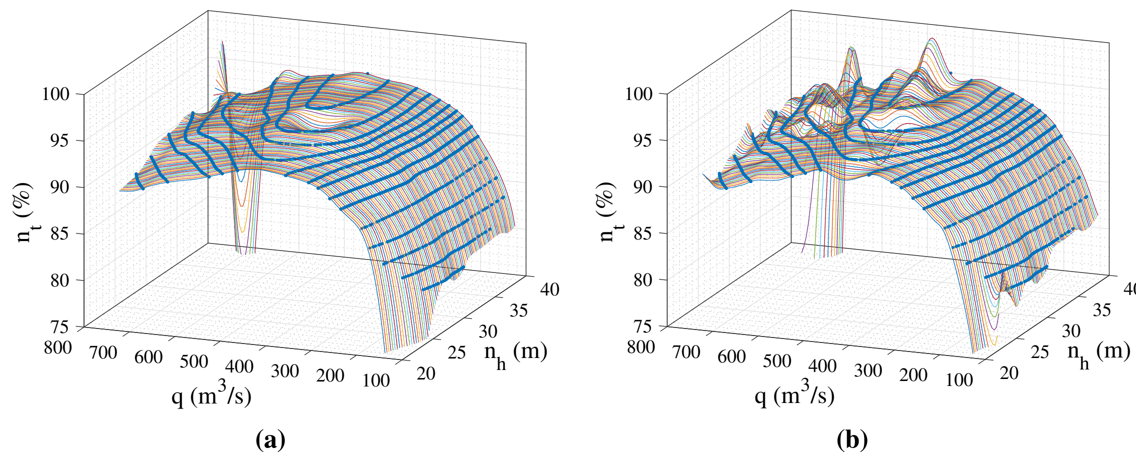


Fig. 4 Distortions on single ANFIS fit with $N_E = 1000$ and **a** $N_F = 7$ and **b** $N_F = 10$

have shown not to have a solid pattern concerning their viability with higher N_E.

Distortions beyond CL first appeared with N_F equal to 4, and further appearances happened unpredictably. For N_F equal to 5 and from 7 to 10, GSs were impracticable for every N_E.

More nets with higher N_E were trained for N_F from 2 to 6 so that this parameter could be further investigated. Table 5 lists the intrinsic errors and shows that the inconsistencies regarding e_{mx} and GS's viability remained.

Regarding the clustered data fitting, the clustered polynomial, ANFIS and ELANFIS fittings will from now on be referred to as FCM-polynomial, FCM-ANFIS and FCM-ELANFIS, respectively.

Table 6 lists the errors for polynomials of degree 5. Since degrees 2–4 presented inferior overall performance, their errors were not exposed. Though most cases granted unviable GSs, E_D clusters granted more viable ones than M_D for degree 2–4 polynomials. The opposite is true for degree 5. Distortions beyond CL occurred only in degrees 4 and 5 fittings and more often in E_D clusters cases. Increasing N_C resulted in errors reduction for almost all cases, however, in many, caused critical discontinuities on the GSs due to regions in space considerably lacking data, as seen between the 94.5% and the 95% n_t levels in Fig. 1a.

Since N_F from 7 to 10 caused surface distortions in the single ANFIS method, N_F from 2 to 6 was tested in the FCM-ANFIS fitting. N_E equal to 1000 was utilized since this parameter's further increase resulted in minor errors reduction in the single ANFIS approach (Table 3), and also due to the significant duration of the simulations. It is important to mention that if one verifies that this method proves to be promising or superior to others in one's problem, one can perform more experiments with different N_E in order to analyze this parameter's influence on errors and GS's integrity. Table 7 lists the errors for this approach. For N_F from 4 to 6, all GSs

Table 3 Single ANFIS errors with increased N_E

N_E	$N_F = 2$		$N_F = 3$		$N_F = 4$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
1500	0.1447′	2.0320′	0.1052′	1.2619′	0.1327′	1.9257′
2000	0.1447′	2.0316′	0.1051′	1.2592′	0.1326′	1.9258′
2500	0.1447′	2.0312′	0.1049′	1.2550′	0.1326′	1.9258′
3000	0.1446′	2.0308′	0.1049′	1.2535′	0.1326′	1.9235′
3500	0.1446′	2.0304′	0.1048′	1.2508′	0.1326′	1.9235′
4000	0.1446′	2.0300′	0.1047′	1.2468′	0.1326′	1.9235′
4500	0.1446′	2.0296′	0.1046′	1.2451′	0.1326′	1.9235′
5000	0.1446′	2.0292′	0.1045′	1.2422′	0.1326′	1.9235′
5500	0.1446′	2.0288′	0.1045′	1.2383′	0.1326′	1.9235′
6000	0.1445′	2.0284′	0.1044′	1.2364′	0.1326′	1.9235′
N_E	$N_F = 5$		$N_F = 6$		–	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	–	–
1500	0.0833	1.3662	0.0567	1.0690	–	–
2000	0.0833	1.3662	0.0567	1.0690	–	–
2500	0.0833	1.3662	0.0567	1.0690	–	–
3000	0.0833	1.3661	0.0567	1.0690	–	–
3500	0.0833	1.3662	0.0567	1.0690	–	–
4000	0.0833	1.3662	0.0567	1.0690	–	–
4500	0.0833	1.3662	0.0567	1.0690	–	–
5000	0.0833	1.3662	0.0567	1.0690	–	–
5500	0.0833	1.3662	0.0567	1.0690	–	–
6000	0.0833	1.3662	0.0567	1.0690	–	–

presented minor to highly severe, though always compromising, distortions. Therefore, it was decided not to expose these cases' errors. An overall tendency to errors reduction can be noticed in Table 7 with higher N_F and N_C, though not so consistently to e_{mx}. As in the FCM-polynomial method, and for

Table 4 Single ELANFIS errors

N_F	$N_E = 100$		$N_E = 200$		$N_E = 500$		$N_E = 1000$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
2	0.1839′	1.3024′	0.1324′	1.4819′	0.1558′	1.4239′	0.1477′	1.5088′
3	0.1086	0.9526	0.1039′	1.2575′	0.0908′	1.1921′	0.0988′	1.3460′
4	0.0860′	1.1879′	0.0893′	1.0962′	0.0922	1.6173	0.0866′	1.0716′
5	0.0725	1.1006	0.0805	0.9177	0.0718	1.1557	0.0663	0.8852
6	0.0669	0.9132	0.0677	1.0203	0.0554	1.1251	0.0529′	0.9825′
7	0.0526	0.8969	0.0480	0.8196	0.0485	0.8999	0.0495	0.8951
8	0.0413	0.9415	0.0387	0.5812	0.0458	0.7956	0.0449	0.6279
9	0.0429	0.6799	0.0410	0.6697	0.0319	0.7821	0.0305	0.7939
10	0.0345	0.5449	0.0279	1.0356	0.0253	0.6316	0.0286	0.7204

Table 5 Single ELANFIS errors with increased N_E

N_E	$N_F = 2$		$N_F = 3$		$N_F = 4$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
1500	0.1425′	1.3623′	0.0959′	1.4838′	0.0865′	1.0540′
2000	0.1387′	1.4979′	0.0986′	0.9556′	0.0792′	0.9092′
2500	0.1384′	1.4261′	0.0966′	1.0299′	0.0828′	1.0434′
3000	0.1495′	1.3680′	0.1017′	0.8676′	0.0845	0.8909
3500	0.1429′	1.5065′	0.0930′	1.0608′	0.0836′	0.8248′
4000	0.1318′	1.3878′	0.0970′	1.3602″	0.0824′	1.1961′
4500	0.1390′	1.3853′	0.0983′	0.8654′	0.0826′	0.7560′
5000	0.1326′	1.3687′	0.0937	0.9997	0.0800′	1.1441′
5500	0.1365′	1.3516′	0.0968′	1.0401′	0.0744′	1.3293′
6000	0.1406′	1.3813′	0.0947′	1.1437′	0.0708′	1.4184′
N_E	$N_F = 5$		$N_F = 6$		–	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	–	–
1500	0.0706′	0.9568′	0.0549	0.9488	–	–
2000	0.0609′	1.2264′	0.0583	0.7468	–	–
2500	0.0617	1.1984	0.0568	0.7321	–	–
3000	0.0678	0.9294	0.0532	0.7868	–	–
3500	0.0691′	0.8786′	0.0609	0.8644	–	–
4000	0.0747′	0.8540′	0.0501	1.1796	–	–
4500	0.0612′	1.2048′	0.0508	1.0469	–	–
5000	0.0629	0.9722	0.0599	0.6500	–	–
5500	0.0647′	1.0999′	0.0574′	0.7807′	–	–
6000	0.0697′	0.9423′	0.0554′	0.9484′	–	–

Table 6 Polynomials of degree 5 errors with FCM clustered data

N_C	E_D		M_D	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}
2	0.0932	1.3228	0.1122′	1.3374′
3	0.1107	1.3398	0.1249′	1.3660′
4	0.0875	1.3418	0.1153	1.5295
5	0.0783	1.1398	0.0989′	1.3653′
6	0.0757	1.1836	0.0929′	1.2059′
7	0.0690′	1.1395′	0.0877	1.1392
8	0.0668	1.1395	0.0823	1.1876
9	0.0653	1.1395	0.0817′	1.0701′
10	0.0598	1.1950	0.0767	0.9446
11	0.0560	0.8996	0.0737	1.1881
12	0.0523	0.9160	0.0691	1.1011

the same reason, some cases with high N_C suffered critical discontinuities.

Analogously to the FCM-ANFIS, only N_F from 2 to 6 was tested in the FCM-ELANFIS fitting. Since changes in N_E were inconclusive in the single ELANFIS method, N_E equal to 1000 was used only that the conditions were the same as the FCM-ANFIS. As before, if one verifies that this method is promising or superior to others in one's problem, one can perform more experiments with different N_E in order to obtain better nets. Table 8 lists this approach's intrinsic errors. All GSs were unviable for N_F from 4 to 6, so their errors were suppressed. Just as in the FCM-ANFIS, higher N_F and N_C provided lower errors, also with less consistency to e_{mx}, and high N_C cases suffered critical discontinuities due to the same motives.

Table 9 lists the errors of the best viable cases of each method. Single polynomial fittings guaranteed continuous surfaces. However, as expected, they provided the highest errors. Single ANFIS granted good surfaces in proper conditions and greatly reduced the errors when compared to the single polynomial method. Although the viability of surfaces originated from single ELANFIS was unpredictable with N_E variation, it was possible to obtain a viable GS that has shown to be superior to the single ANFIS best one. Regarding the methods applied to FCM clustered data, although most cases were unviable, the best viable one of each method showed to be superior to the best one of the respective method fitting all

Table 7 ANFIS errors with FCM clustered data

N_C	E_D				M_D			
	$N_F = 2$		$N_F = 3$		$N_F = 2$		$N_F = 3$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
2	0.1112	1.4796	0.0605′	1.2781′	0.1488	1.8015	0.0878′	1.1311′
3	0.1174	1.4541	0.0713′	1.2661′	0.1530	2.0317	0.1007′	1.2803′
4	0.0945	1.3997	0.0590	1.2896	0.1316′	1.6934′	0.0827	1.4107
5	0.0786	1.1004	0.0414	0.4156	0.1206	1.2438	0.0741′	1.0033′
6	0.0705′	1.2687′	0.0494	0.7858	0.1168	1.3024	0.0605	0.8934
7	0.0599′	1.1029′	0.0338	0.4374	0.0997	1.2555	0.0583	0.9459
8	0.0580′	1.1032′	0.0321	0.4450	0.0962	1.2095	0.0500	0.9815
9	0.0558′	1.1003′	0.0307	0.4295	0.0853′	1.0992′	0.0537	0.7700
10	0.0493	1.0221	0.0246	0.4342	0.0884	1.1128	0.0454	0.8263
11	0.0519′	0.9081′	0.0235	0.7743	0.0836	1.4128	0.0440	1.1380
12	0.0459	0.7498	0.0199	0.7741	0.0751	1.5885	0.0410	1.0325

Table 8 ELANFIS errors with FCM clustered data

N_C	E_D				M_D			
	$N_F = 2$		$N_F = 3$		$N_F = 2$		$N_F = 3$	
	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}	e_{mn}	e_{mx}
2	0.1038	1.1615	0.0651	1.2644	0.1250′	1.5916′	0.0811′	1.3376′
3	0.1211	1.2780	0.0829	1.4306	0.1446	1.5928	0.0889	0.8613
4	0.0876	1.2250	0.0704	0.9977	0.1196	1.4398	0.0826	0.9787
5	0.0885	1.1290	0.0492	1.0960	0.1051	1.3923	0.0732′	0.9006′
6	0.0734′	1.3036′	0.0494′	0.9448′	0.1043	1.1810	0.0681′	1.2406′
7	0.0724′	1.2117′	0.0428	1.2576	0.0900	1.5013	0.0573	0.9034
8	0.0687′	1.3032′	0.0408	0.7293	0.0857	1.2236	0.0586	0.8946
9	0.0682	1.2610	0.0424	0.8936	0.0763	1.5605	0.0500	1.0002
10	0.0634	1.2214	0.0340	1.0506	0.0752	1.0067	0.0432	1.0282
11	0.0602	0.7119	0.0308	0.7263	0.0763	1.2684	0.0498	0.8925
12	0.0513	1.0007	0.0267	0.6665	0.0664	1.2290	0.0436	0.7377

data at once, including the polynomials of degrees 2, 3 and 4, which did not have their errors exposed in this paper. Single ANFIS and ELANFIS provided viable surfaces for N_F from 4 to 6, though their clustered versions did not. This fact indicates that increased N_F requires more samples in order to guarantee satisfactory surfaces.

In the particular problem of TEC fitting, the CL delimitates an area where distortions are not relevant. If this characteristic had been left aside and the whole curve required a smooth and well-behaved fit, some of the best cases would not be the ones in Table 9. Table 10 lists the errors of the best cases for the methods that changed in this scenario. Either way, three membership functions ELANFIS fitting 6 E_D clusters provided the best result. Figure 5 shows this case's GS.

Results indicated that E_D clustering grants better fittings concerning not only errors but also GS's integrity, at least for this specific problem. For the current phase of our project, the FCM-ELANFIS method was chosen as the definitive method to fit the TEC. In order to search for a definitive net, more simulations with diverse N_E and promising values of N_C and N_F will be performed.

Table 9 All methods' best viable cases' errors

Method	Parameters	e_{mn}	e_{mx}
Single polynomial	Degree 5	0.1424	1.3786
Single ANFIS	$N_F = 6$; $N_E = 200$	0.0869	1.2096
Single ELANFIS	$N_F = 6$; $N_E = 1000$	0.0529	0.9825
FCM-polynomial	$N_C = 7$; E_D; Degree 5	0.0690	1.1395
FCM-ANFIS	$N_C = 11$; E_D; $N_F = 2$	0.0519	0.9081
FCM-ELANFIS	$N_C = 6$; E_D; $N_F = 3$	0.0494	0.9448

Table 10 Best viable cases' errors assuming CL does not exist

Method	Parameters	e_{mn}	e_{mx}
Single ANFIS	$N_F = 3$; $N_E = 6000$	0.1044	1.2364
Single ELANFIS	$N_F = 6$; $N_E = 6000$	0.0554	0.9484
FCM-polynomial	$N_C = 6$; M_D; Degree 5	0.0929	1.2059

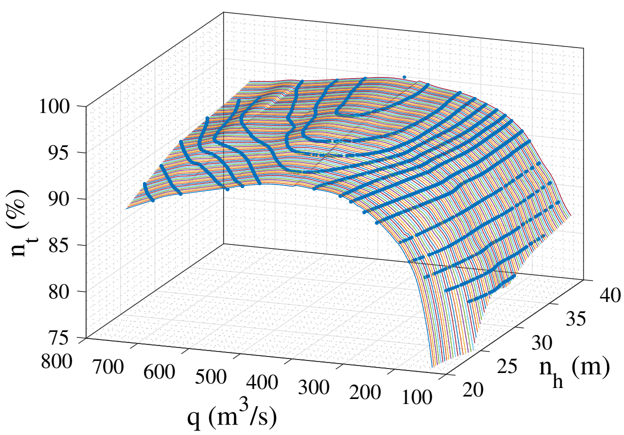


Fig. 5 Best fitting achieved—FCM-ELANFIS with $N_C = 6$, E_D and $N_F = 3$

5 Conclusion

This paper presented a study of Fuzzy C-Means, ANFIS and Extreme Learning ANFIS applied to the fitting of a hydroelectric power plant's turbine efficiency curve, which is a nonlinear three-dimensional curve. The adjustment performed by the Extreme Learning ANFIS applied to a clustered dataset granted the best results. The analysis led to the conclusion that depending on the data distribution, ANFIS and Extreme Learning ANFIS may provide great surfaces when in combination with Fuzzy C-Means clustering.

As possible further works, it is intended to study the application of other overlapping clustering algorithms in the presented problem: for instance, the ones mentioned in Sect. 1, especially those based on type-2 fuzzy. It is also planned to test other types of distance measurement when clustering the dataset, and, regardless of incorporating clustering or not, to apply other artificial intelligence-based techniques to the fitting problem, such as deep learning, which has shown to be successful in several problems.

Acknowledgements The authors thank the support of the Electrical Engineering Post Graduation Program (PPEE) of the Federal University of Juiz de Fora (UFJF), INERGE and EDP. The development presented in this paper came from the ANEEL R&D project titled "Integrated Water, Cost and Energy Management Using Stochastic Models in Series with Deep Learning to Increase the Efficiency of Hydroelectric Power Plants - Rain Flow" (PD-00673-0052/2018). The authors thank ANEEL and the technicians of all the companies involved.

References

1. Alam I, Ansari MA, Pal NS (2016) A comparative study between wavelet primarily based ANN and ANFIS algorithm technique to locate fault in a transmission line. In: 2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES). IEEE, pp 1–6
2. Arce A, Ohishi T, Soares S (2002) Optimal dispatch of generating units of the itaipú hydroelectric plant. IEEE Trans Power Syst 17(1):154–158
3. Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10(2–3):191–203
4. Bortoni EC, Bastos GS, Souza LE (2007) Optimal load distribution between units in a power plant. ISA Trans 46(4):533–539
5. Cao H, Si G, Zhang Y, Ma X (2007) A hybrid controller of self-optimizing algorithm and anfis for ball mill pulverizing system. In: 2007 international conference on mechatronics and automation. IEEE, pp 3289–3294
6. Cao J, Lin Z, Huang GB, Liu N (2012) Voting based extreme learning machine. Inf Sci 185(1):66–77
7. Cao JJ, Kwong S, Wang R, Li K (2012) A weighted voting method using minimum square error based on extreme learning machine. In: 2012 international conference on machine learning and cybernetics, vol 1. IEEE, pp 411–414
8. Civicioglu P (2007) Using uncorrupted neighborhoods of the pixels for impulsive noise suppression with anfis. IEEE Trans Image Process 16(3):759–773
9. Cordova M, Finardi E, Ribas FAC, de Matos V, Scuzziato M (2014) Performance evaluation and energy production optimization in the real-time operation of hydropower plants. Electr Power Syst Res 116:201–207
10. Daneshi H, Srivastava AK, Daneshi A (2010) Application of fuzzy logic to priced-based unit commitment under price uncertainty. In: IEEE PES T&D 2010. IEEE, pp 1–5
11. Dias BH, Tomim MA, Marcato ALM, Ramos TP, Brandi RBS, da Silva Junior IC, Passos Filho JA (2013) Parallel computing applied to the stochastic dynamic programming for long term operation planning of hydrothermal power systems. Eur J Oper Res 229(1):212–222
12. Erol H, Tyoden BM, Erol R (2018) Classification performances of data mining clustering algorithms for remotely sensed multispectral image data. In: 2018 innovations in intelligent systems and applications (INISTA). IEEE, pp 1–4
13. Fernandes JPT, de Barros Correia P, Hidalgo IG, Colnago GR (2013) A genetic algorithm solution for optimization of the power generation potential in hydroelectric plants. In: 2013 IEEE congress on evolutionary computation. IEEE, pp 2504–2511
14. Finardi EC, Scuzziato MR (2013) Hydro unit commitment and loading problem for day-ahead operation planning problem. Int J Electr Power Energy Syst 44(1):7–16
15. Finardi EC, Takigawa FYK, Brito BH (2016) Assessing solution quality and computational performance in the hydro unit commitment problem considering different mathematical programming approaches. Electr Power Syst Res 136:212–222
16. Havens TC, Bezdek JC, Leckie C, Hall LO, Palaniswami M (2012) Fuzzy c-means algorithms for very large data. IEEE Trans Fuzzy Syst 20(6):1130–1146
17. Hidalgo I, Soares FS, Fontane D, Lopes JE, Cicogna M, Silva FD (2012) Analysis of the level-release polynomial from a hydroelectric plant. Syst Cybern Inform 10(1):62–65
18. Hidalgo IG, Fontane DG, Soares FS, Cicogna MA, Lopes JE (2010) Data consolidation from hydroelectric plants. J Energy Eng 136(3):87–94
19. Hidalgo IG, Soares S, Fontane DG, Cicogna MA, Lopes JE (2009) Impact of the data quality from hydroelectric plants in the past oper-

ation analysis using a middle term simulation tool. In: International multi-conference on engineering and technological innovation
20. Hore P, Hall LO, Goldgof DB (2007) Single pass fuzzy c means. In: 2007 IEEE international fuzzy systems conference. IEEE, pp 1–7
21. Hossain MS, Muhammad G (2019) Emotion recognition using deep learning approach from audio-visual emotional big data. Inf Fusion 49:69–78
22. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1–3):489–501
23. Hung MC, Yang DL (2001) An efficient fuzzy c-means clustering algorithm. In: Proceedings 2001 IEEE international conference on data mining. IEEE, pp 225–232
24. Jagtap P, Pillai GN (2014)Comparison of extreme-anfis and anfis networks for regression problems. In: 2014 IEEE international advance computing conference (IACC). IEEE, pp 1190–1194
25. Jang JSR (1993) Anfis: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23(3):665–685
26. Krishnapuram R, Keller JM (1993) A possibilistic approach to clustering. IEEE Trans Fuzzy Syst 1(2):98–110
27. Miyamoto S, Umayahara K (2000) Methods in hard and fuzzy clustering. In: Liu ZQ, Miyamoto S (eds) Soft computing and human-centered machines. Springer, Tokyo, pp 85–129
28. Navin NK, Sharma R (2016) A modified differential evolution approach to PHEV integrated thermal unit commitment problem. In: 2016 IEEE 7th power India international conference (PIICON). IEEE, pp 1–4
29. Nhabangue MFC, Pillai GN, Sharma ML (2018) Chaotic time series prediction with functional link extreme learning anfis (fl-elanfis). In: 2018 international conference on power, instrumentation, control and computing (PICC). IEEE, pp 1–6
30. Pal NR, Pal K, Bezdek JC (1997) A mixed c-means clustering model. In: Proceedings of 6th international fuzzy systems conference, vol 1. IEEE, pp 11–21
31. Pal NR, Pal K, Keller JM, Bezdek JC (2005) A possibilistic fuzzy c-means clustering algorithm. IEEE Trans Fuzzy Syst 13(4):517–530
32. Parker JK, Hall LO (2013) Accelerating fuzzy-c means using an estimated subsample size. IEEE Trans Fuzzy Syst 22(5):1229–1244
33. Pillai GN (2014) Extreme anfis: a new learning machine for faster learning. In: Presented in international conference on knowledge collaboration in engineering, held at Coimbatore
34. Raffe W, Hu J, Zambetta F, Xi K (2010) A dual-layer clustering scheme for real-time identification of plagiarized massive multiplayer games (mmg) assets. In: 2010 5th IEEE conference on industrial electronics and applications. IEEE, pp 307–312
35. Ramirez-Rozo TJ, Garcia-Alvarez JC, Castellanos-Dominguez CG (2012) Infrared thermal image segmentation using expectation-maximization-based clustering. In: 2012 XVII symposium of image, signal processing, and artificial vision (STSIVA). IEEE, pp 223–226
36. Rhee FCH, Hwang C (2001) A type-2 fuzzy c-means clustering algorithm. In: Proceedings joint 9th IFSA world congress and 20th NAFIPS international conference (Cat. No. 01TH8569), vol 4. IEEE, pp 1926–1929
37. Ribbens A, Hermans J, Maes F, Vandermeulen D, Suetens P (2013) Unsupervised segmentation, clustering, and groupwise registration of heterogeneous populations of brain MR images. IEEE Trans Med Imaging 33(2):201–224
38. Rubio E, Castillo O, Melin P (2015) A new interval type-2 fuzzy possibilistic c-means clustering algorithm. In: 2015 annual conference of the North American fuzzy information processing society (NAFIPS) held jointly with 2015 5th world conference on soft computing (WConSC). IEEE, pp 1–5
39. Shihabudheen KV, Pillai GN (2015) Internal model control based on extreme learning ANFIS for nonlinear application. In: 2015 IEEE international conference on signal processing, informatics, communication and energy systems (SPICES). IEEE, pp 1–5
40. Silva IC Jr, Carneiro S Jr, de Oliveira EJ, Pereira J, Garcia PA, Marcato AL (2008) A Lagrangian multiplier based sensitive index to determine the unit commitment of thermal units. Int J Electr Power Energy Syst 30(9):504–510
41. Sönmez Y, Tuncer T, Gökal H, Avcı E (2018) Phishing web sites features classification based on extreme learning machine. In: 2018 6th international symposium on digital forensic and security (ISDFS). IEEE, pp 1–5
42. Terry LA, Pereira MVF, Araripe Neto TA, Silva LFCA, Sales PRH (1986) Coordinating the energy generation of the Brazilian national hydrothermal electrical generating system. Interfaces 16(1):16–38
43. Tsai CF, Huang SC (2015) An effective and efficient grid-based data clustering algorithm using intuitive neighbor relationship for data mining. In: 2015 international conference on machine learning and cybernetics (ICMLC), vol 2. IEEE, pp 478–483
44. Tushar A, Pillai G, et al (2015) Extreme learning anfis for classification problems. In: 2015 1st international conference on next generation computing technologies (NGCT). IEEE, pp 784–787
45. Xie W, Xie Z, Zhao F, Ren B (2018) POLSAR image classification via clustering-WAE classification model. IEEE Access 6:40041–40049
46. Yang MS, Ko CH (1996) On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets Syst 84(1):49–60
47. Zarandi MHF, Zarinbal M, Türksen IB (2009) Type-ii fuzzy possibilistic c-mean clustering. In: IFSA/EUSFLAT conference. Citeseer, pp 30–35
48. Zheng R, Pan F (2014) Soft sensor of glutamate concentration using extreme learning machine. In: Proceeding of the 11th world congress on intelligent control and automation. IEEE, pp 1865–1868