International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems

Checking Orthogonal Transformations and Genetic Algorithms for Selection of Fuzzy Rules based on Interpretability-Accuracy Concepts

M. Isabel Rey

INDOMAUT S.L.
Pol. Ind. San Cristóbal
47012 Valladolid, SPAIN
Email: i.rey@indomaut.com

Marta Galende

CARTIF Centro Tecnológico ⋆
47151 Boecillo (Valladolid), SPAIN
Email: margal@cartif.es

M. J. Fuente

Dpt. of Systems Engineering and Control
School of Industrial Engineering
University of Valladolid
47011 Valladolid, SPAIN
Email: maria@autom.uva.es

Gregorio I. Sainz-Palmero [†,⋆]

Dpt. of Systems Engineering and Control †
School of Industrial Engineering
University of Valladolid
47011 Valladolid, SPAIN
Email: gresai@{cartif.es, eii.uva.es}

Received (received date)
Revised (revised date)

Fuzzy modeling is one of the most known and used techniques in different areas to model the behavior of systems and processes. In most cases, as in data-driven fuzzy modeling, these fuzzy models reach a high performance from the point of view of accuracy, but from other points of view, such as complexity or interpretability, they can present a poor performance.

Several approaches are found in the bibliography to reduce the complexity and improve the interpretability of the fuzzy models. In this paper, a post-processing approach is carried out via rule selection, whose aim is to choose the most relevant rules for working together on the well-known accuracy-interpretability trade-off. The rule relevancy is based on Orthogonal Transformations, such as the SVD-QR rank revealing approach, the P-QR and OLS transformations. Rule selection is carried out using a genetic algorithm that takes into account the information obtained by the Orthogonal Transformations. The main objective is to check the true significance, drawbacks and advantages of the rule selection based on the orthogonal transformations via the rule firing strength matrix.

In order to carry out this aim, a neuro-fuzzy system, FasArt (Fuzzy Adaptive System ART based), and several case studies, data sets from the KEEL Project Repository, are used to tune and check this selection of rules based on orthogonal transformations, genetic selection and accuracy-interpretability trade-off. This neuro-fuzzy system generates Mamdani fuzzy rule based systems (FRBSs), in an approximative way. NSGA-II is the MOEA tool used to tune the proposed rule selection.

Keywords: Fuzzy Systems, Interpretability, Accuracy, Rule Selection, Orthogonal Transformations, Genetic Algorithm

1. Introduction

Fuzzy modeling is one of the most known approaches for a wide range of problems. Data-driven rule based fuzzy models have been used in several and very different scientific and technical areas [1,2,3,4,5].

In general, most of the fuzzy models taken into consideration in real world applications have been data-driven and rule based fuzzy models due to their advantages: easy use and performance. This performance has usually been evaluated on the basis of the accuracy of the models, thus minimizing the error between the real and the estimated output generated by the fuzzy models. But other aspects have not been taken into consideration: complexity, interpretability, etc. Some of them are basic principles of fuzzy logic, but data-driven fuzzy models use them as simple mathematical tools, losing their original fuzzy meaning.

Complexity is a very usual index or measure, and it is a problem in data-driven rule based fuzzy models related with other aspects of these models, such as their accuracy and interpretability. Thus, if a reduction of this complexity was reached, it could permit a better performance of these other aspects to be reached, so as to obtain better fuzzy models. The question is the way in which this complexity reduction or model improvement can be carried out. Different approaches to this question can be found in [6,7,8,9,10].

In this work, the complexity reduction is studied based on Orthogonal Transformations and accuracy-interpretability trade-off by a genetic rule selection. Orthogonal transformations [11,12] have been one of the alternative approaches for complexity reduction and interpretability improving of fuzzy models [9,13,14,15,16,17,18,19]. This approach is focused on orthogonal transformations applied on the *firing strength matrix* of the fuzzy model rules as a regression problem, in order to estimate the relevance of the rules, then a rule selection is carried out. Each transformation has its own strategy. In [14,20] some general comments and ideas about the research of these transformations can be found, but in general, there is not very much extensive and far-reaching experimentation that clearly specifies the criteria for carrying out the rule selection.

In this context, this work checks the possibilities and drawbacks of the orthogonal transformations as a postprocessing approach to simplify and get more interpretable approximative rule-based fuzzy models. Thus, an approximative fuzzy model is considered involving: accuracy-interpretability criteria to give relief in some of the orthogonal transformation weaknesses, rule selection by a genetic approach subject to constraints on influential rules and accuracy-interpretability trade-off and checking the influence or relevance of the rules selected. On other hand, this can give ideas and support to define some criteria in order to pick up the best rules of the fuzzy models so as to get more simple, accurate and interpretable data-driven fuzzy models.

The paper is organized as follows: first, in Section 2, a brief description of alternative points of view about fuzzy modeling, interpretability and accuracy are given. This section includes a description of the main concepts of orthogonal transformations. The proposal of genetic rule selection based on orthogonal transformations and accuracy-interpretability trade-off is introduced in Section 3. In Section 4, the methodology used in this work is described. Some experimental studies are carried out and the main results obtained are discussed in Section 5. Finally, in Section 6, the most interesting conclusions obtained from this work are set out.

2. Fuzzy Modeling: Accuracy vs. Interpretability

Initially, two well known modeling approaches to generate fuzzy rules are described in the bibliography [21,22,23]:

(1) *Precise Fuzzy Modeling*, whose main goal is to obtain a model which is as accurate as possible. In general, the models generated have a good accuracy but a low level of interpretability. This modeling is popular with data-driven knowledge but expert knowledge is also considered.
(2) *Linguistic Fuzzy Modeling*, these models have a good level of interpretability but poor accuracy. Here, knowledge from experts and data guide the modeling process.

Both approaches have their own drawbacks and advantages, but there are several ways to deal with the generation of fuzzy systems whose performance includes an adequate accuracy-interpretability trade-off. This trade-off question is an open one: in what way are fuzzy systems more interpretable and accurate enough?. Some reviews of interpretability and the way in which this can be achieved can be found in [8,10,20]. Sometimes, these appear associated with the concepts of complexity and explanation capability [24], which can be considered as indirect measures to evaluate the interpretability. In some works, for instance [14,15,16,25,26], the reduction of the complexity system can imply a better interpretability of the fuzzy system. In any case, the interpretability of fuzzy systems is still a point of discussion amongst researchers [7,27].

One of the above-mentioned approaches, perhaps not the most popular, is based on orthogonal transformations applied on the firing strength matrix of the fuzzy model rules. The goal is to estimate the most influential rules, which are selected, so the interpretability is improved by reducing the complexity, in this case by reducing the number of rules [9,14,16,28]. Several orthogonal transformations are taken into consideration in this research domain, some of the most popular are: *Singular Value Decomposition (SVD)*, *Pivoted QR(P-QR)* and *Orthogonal Least-Squares (OLS)* decompositions that are considered in this work.

2.1. *Orthogonal Transformations and Complexity Reduction*

The orthogonal transformations are used for rule selection/reduction in fuzzy modeling in two main approaches [11,14]:

- *Rank-revealing approach,* an estimation rank of the firing strength matrix is given. In this approach the *SVD and SVD-QR Decompositions* are considered.
- *Rule subset selection,* the individual contributions of the rules are evaluated to reach their *importance ordering*[16]. The *Pivoted QR(P-QR)* and *Orthogonal Least-Squares (OLS)* transformations are included in this approach.

In this context, a fuzzy model can be written as a linear regression problem (Eq. 1) [14,16]:

$$y = P * \theta + e \tag{1}$$

where: $y = [y_1, y_2 ..., y_N]^T$ are the measured outputs, $\theta = [c_1, c_2, ..., c_M]^T$ are the consequents of the M rules and $e = [e_1, e_2, ..., e_N]^T$ are the vector of approximation errors. The matrix $P = [p_1, p_2, ..., p_M] \in R^{N \times M}$ contains the firing strength of all the M rules for the N inputs x_k, where $p_i = [p_{i1}, p_{i2}, ..., p_{iN}]^T$

- **SVD Decomposition** is used to determine the effective rank of the rule firing matrix (P). This can be expressed as (Eq. 2):

$$p_i(x) = \frac{\prod_{j=1}^{N} A_{ij}(x_j)}{\sum_{k=1}^{M} \prod_{j=1}^{N} A_{kj}(x_j)} \tag{2}$$

where $x = [x_1, .., x_N]^T$ is the input vector, $A_{i1}, ...A_{iN}$ are fuzzy sets defined in the antecedent space and M is the number of rules of the fuzzy model.
The pseudoinverse of P is obtained from the singular value decomposition (SVD) of P (Eq. 3):

$$P = U \Sigma V^T \tag{3}$$

where $U \in R^{N \times M}$ and $V \in R^{N \times M}$ are orthogonal matrices and $\Sigma \in R^{M \times M}$ is a diagonal matrix with the singular values: $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq ... \geq \sigma_M \geq 0$ in decreasing order as diagonal. The pseudoinverse is (Eq. 4):

$$P^+ = V \Sigma^+ U^T \tag{4}$$

where $\Sigma^+ \in R^{M \times M}$ is a diagonal matrix with the reciprocals $1/\sigma_1, 1/\sigma_2 .....1/\sigma_r$. The number of nonzero singular values in the SVD of P reveals the rank of P. The rank estimation, r, is not evident if a "*gap*" to discriminate singular values is not identified, which is not usual. The most important rules are those associated with higher singular values.

- **P-QR Decomposition**, this approach can produce a rule ordering without a rank estimation. Here, P-QR is directly applied to P, obtaining a permutation matrix [11]: The QR decomposition of P is given by $P * \Pi = Q * R$, where $\Pi \in \Re^{M*M}$ is a permutation matrix, $Q \in \Re^{N*M}$ has orthogonal columns and $R \in \Re^{M*M}$ is upper triangular (Eq. 5), such that

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{kk} \end{bmatrix} \tag{5}$$

  The diagonal values of R are called *R-values* ($|R_{kk}|$) [9], which track the singular values $\sigma(P)$, so the most active and least redundant rules are those whose R-values are higher [16] in the original fuzzy rule space.
- **OLS Decomposition**: here, the firing matrix P is decomposed into a set of orthogonal vectors to evaluate the individual contribution of each rule: $P = WA$, where W is an orthogonal matrix such as $W^T W = I$, and A is an upper-triangular matrix with unity diagonal values. Then, substituting $P = WA$ into (Eq. 1), we have $y = WA\theta + e = Wg + e$, where $g = A\theta$. Since the colums w_i of W are orthogonals the sum of squares of y can be written as (Eq. 6):

$$y^t y = \textstyle\sum_{i=1}^{M} g_i w_i^T w_i + e^T e \tag{6}$$

  The part of the output variance $y^t y/N$ described by the regressors is $\sum_i w_i^T w_i/N$. Then an error reduction due to the rule i is: $[err] = \frac{g_i^2 w_i^T w_i}{y^t y}$. Thus, this can be used to define an importance ordering for the rules and to carry out a selection.

In short, some comments on these transformations can be made in the domain of rule reduction [14]: a) the *SVD* and *P-QR* transformations do not pay attention to the output contribution of the rule, b) rank-revealing methods are conservative in the rule reduction due to the difficulty of estimating the rank of P, c) *OLS Decomposition* does not consider the structure of the rules in terms of redundancy, similarity, etc.

In order to avoid these weaknesses in this work, concepts about interpretability, accuracy and their trade-off are used together with the orthogonal transformations in order to address the rule selection.

3. Genetic Rule Selection Based on Orthogonal Transformations and Accuracy-Interpretability trade-off

The main objective of this work is to check the advantages and drawbacks of the orthogonal transformations for rule ordering and selection. Three well-known orthogonal transformations have been involved in this work: SVD, P-QR and OLS.

The fuzzy models are generated by a neuro-fuzzy system, FasArt [29,30], which is considered an approximative fuzzy system that is very popular in engineering domains. Now, in order to generate a better rule selection, this is carried out following the guidance of different points of views concerned with the:

- Relevance or influence of each fuzzy rule defined by the orthogonal transformations.
- Accuracy-Interpretability trade-off in fuzzy models defined by measures on both concepts.

These aspects are complementary and try to avoid some of the drawbacks of orthogonal transformations:

- SVD and P-QR only consider the rule antecedents, so the accuracy of the model can give support in both cases.
- OLS does not manage well the redundancy and similarity of the rules, so the interpretability index can mitigate this problem.
- The importance ordering for rules provides an individual evaluation for the rule selection.
- On the other hand, the accuracy-interpretability trade-off gives a global index of the quality of the rule selection carried out.

In order to check all this, a genetic approach for the rule selection is done. This provides an interesting scenario of results concerning the rule ordering and selection based on these orthogonal transformations. The study of this scenario will give us a better knowledge of the scope of this selection proposal, and will give response and support to some open questions about this type of ordering and selection, such as:

- How much influence must be preserved by the selected rules?
- How many rules must be selected?
- What is the role of the rules with lower relevance by orthogonal transformations?
- Must these lower relevance rules be considered in order to achieve good models?
- etc.

In the following subsections, a brief description of the accuracy-interpretability measures considered is done. Then, some comments and references on the genetic and neuro-fuzzy approach used in this work are introduced.

3.1. *Accuracy and Interpretability Measures*

The accuracy and interpretability measures considered in this work are defined in [31]. The *accuracy* of the model is measured through its Mean Squared Error (MSE) (Eq. 7):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i')^2 \tag{7}$$

The *interpretability* measure is an aggregating index based on similarity and complexity ideas. In both cases, a lower value has a positive influence on reducing the complexity and improving the interpretability of the fuzzy models. These measures about similarity and complexity are:

- *Compactness* or *Number of rules* (RN).
- *Similarity* amongst rules (S).
- *Redundancy* of the fuzzy rule set (R).
- *Incoherency* of rules (I).
- *Completeness* or *No-Coverage* (C).

Thus, the aggregation index to measure the interpretability is formulated as follows (Eq. 8):

$$\begin{aligned} &Inter_C = ArithmeticMean(\\ &\quad \lambda_{nr} * RuleNumber_{nor}, \lambda_s * Similarity_{nor}, \\ &\quad \lambda_r * Redundancy_{nor}, \lambda_i * Incoherency_{nor}, \\ &\quad \lambda_{nc} * NoCoverage_{nor}) \\ &\lambda_j \in (0,1) \end{aligned} \tag{8}$$

Here $\lambda_j = 1\ \forall j$, and the normalization is (Eq. 9):

$$Index_{nor} = 1 - \tfrac{Index_{Original} - Index_{Current}}{Index_{Original}} \tag{9}$$

3.2. *Genetic Algorithms and Neuro-Fuzzy Systems*

3.2.1. *Genetic Algorithms*

Genetic algorithms, genetic programming, and evolutionary strategies, among other evolutionary algorithms (EAs), are very popular tools to tune fuzzy models [32,33]. A general taxonomy of this is introduced in [21] where genetic algorithms are used in two alternative ways to generate fuzzy systems: tuning and learning. There are some papers and contributions that use multi-objective evolutionary algorithms (MOEAs) to improve the accuracy-interpretability trade-off by taking into account these two ways [a].

Within this taxonomy, an alternative is to use MOEAs to select a subset of cooperative rules from a set of candidate fuzzy rules. Then, the objective is to

[a] A list of papers on this domain can be found at http://www.iet.unipi.it/m.cococcioni/emofrbss.html

obtain a more reduced rule set, improving its original performance, usually the accuracy and the interpretability.

The well-known multi-objective evolutionary algorithm, NSGA-II [34], is taken into account in this work, but other multi-objetive evolutionary algorithms can also be used [35]. Two fitness functions from MSE (Eq. 7) and $Inter_C$ (Eq. 8) are used to get a fuzzy model with better accuracy-interpretability trade-off.

A third fitness function is used to penalize lower values from importance ordering generated by orthogonal transformations. According to previous works [14,16], these rules introduce a high level of similarity, low level of activity and high redundancy, so they must be avoided. Thus, this is implemented as follows (Eq. 10):

$$Penalty_{Singular\ Value/R\text{-}value/Variance} = \sqrt[n]{\prod_{i=1}^{n}\left(1-(SingularValue/R-value/Variance)_{norm_i}\right)} \tag{10}$$

where:

$$(Singular\ Value/R\text{-}value/Variance)_{norm_i} = \frac{(Singular\ Value/R\text{-}value/Variance)_{Rule_i}}{\sum_{j=1}^{n}(Singular\ Value/R\text{-}value/Variance)_{Rule_j}} \tag{11}$$

3.2.2. *Neuro-Fuzzy System FasArt*

On the other hand, neuro-fuzzy systems are a very popular approach to generate FRBSs, taking advantage of the learning capacity of Artificial Neural Networks (ANN) and the explanatory capacity of Fuzzy Logic. In this work, the neuro-fuzzy system FasArt [29,30], which is a neuro-fuzzy system based on the Adaptive Resonance Theory (ART) has been used. FasArt introduces an equivalence between the activation function of each FasArt neuron and a membership function. In this way, FasArt is equivalent to a Mamdani-type FRBS with: Fuzzification by single point, Inference by product, and Defuzzification by average of fuzzy set centers. A full description of this model can be found in [29] and [30]. If the taxonomy for FBRSs described in [21] is taken into account, FasArt is an approximate model. Another classification can be done if [23,36] are considered: FasArt is a Mamdani-type FRBS for precise modeling.

This FasArt system has been used in several previous works [37,38] for modeling, fault detection, pattern recognition, etc, with reasonable results when its accuracy as a fuzzy model is involved; but when other aspects, such as rule interpretability, are considered, then some problems appear; so this system is an adequate instance for checking this proposal. Most of these aspects are common for models based on ART Theory, and they have been treated in different works [39,40].

Table 1. FasArt Parameters for modeling

FasArt Parameters
$\rho_A = \rho_B = 0.3$
$\gamma_A = \gamma_B = 10$

4. Methodology

In this paper, the proposed methodology is focused on checking the capacities of the orthogonal transformations for rule selection based on accuracy-interpretability trade-off and genetic tuning. This goal is reached using a general post-processing fuzzy rule selection through a three-objective genetic approach: accuracy, interpretability and the most influential rules. In this scenario, it will be possible to check the trade-off of the fuzzy models tuned by the rule selection, the rule influence level preserved in the simplified models, the level of complexity reduction achieved, the distribution of the rule influence amongst the selected rules for each model, etc.

The fuzzy models were generated by FasArt in five fold cross validation for each regression problem considered (see data sets in Section 5). The FasArt parameters considered for all the cases are shown in Table 1, where ρ_A=ρ_B is the vigilance parameter used by FasArt and $\gamma_A = \gamma_B$ is the fuzzification rate in FasArt.

A general methodology description is summarized in Algorithm 1. This methodology is set out in the following sections, describing in detail the MOEA applied in the post-processing stage for this rule selection.

4.1. *Multi-Objective Evolutionary Algorithm for Rule Selection*

The fuzzy rule selection to achieve *lower complexity and better performance on interpretability with enough accuracy* based on the influential rules is carried out by a MOEA. In order to achieve the aims commented previously, a three-objective ($Inter_C$, *Acc*, *Penalty*) genetic approach is used based on the well-known NSGA-II algorithm [34].

In the next sections, the fitness functions are formulated and the genetic parameters and operators are described.

4.1.1. *Objectives*

The fitness functions are shown in Eq. (12), here some performance desired for the model can be taken into account.

$$\begin{aligned}
&max(Accuracy) = min(MSE_{tra}) \\
&max(Interpretability) = min(Inter_C) = \\
&\quad = min(AritmeticMean(\lambda_j * InterpretabilityIndex_j)) \\
&min(Penalty_{OTv_s})
\end{aligned} \tag{12}$$

Algorithm 1 Methodology for Genetic Rule Selection based on Accuracy-Interpretability Trade-Off and Orthogonal Transformations

for Neuro-Fuzzy Algorithm=$FasArt$ **do**
  for OT=$SVD - QR{:}P - QR{:}OLS$ **do**
    for $DataSet = 1$ to 9 **do**
      for $CrossValidation = 1$ to 5 **do**
        Generation of Rule Importance Ordering by OT
        Training Neuro-Fuzzy System ($\rho_A = \rho_B = 0.3$ and $\gamma_A = \gamma_B = 10$)
        for $Run = 1$ to 6 **do**
          Generate Initial Population
          Run Genetic Algorithm NSGAII (Selection-Binary Tournament, Crossover-HUX P_c=0.9, Mutation-Classical P_m=0.7, Population size-100 and Evaluations-50000)
        end for
      end for
      Analysis Pareto Front (DataSet) {Best $Inter_C$, Median $Acc - Inter_C$ and Best Acc}
    end for
  end for
end for
Non-Parametric Statistical Test

The three-objective genetic algorithm must get a fuzzy model with better accuracy-interpretability trade-off based on the most influential rules and importance ordering provided by orthogonal transformations:

- Maximizing the accuracy evaluated by Mean Squared Error (MSE) (Eq 7).
- Maximizing the interpretability of the fuzzy model guided by complexity concepts $Inter_C$ defined in Section 3.1 (Eq 8).
- Minimizing the number of selected rules with low influence or importance (Eq 10).

4.1.2. *Coding Scheme, Populations and Genetic Operators*

In order to run NSGA-II, the following characterization is done:

- Individuals are coded by *binary-coding*: $S = s_1 s_2 ... s_N$ (N is the number of initial rules), where $s_q = 0$ shows that the rule R_q is not included, while $s_q = 1$ shows the rule is present.
- Genes take the value 1 for all of the individuals of the *initial population* in order to achieve a progressive extraction of the worst rules.
- *Genetic operators* selected according to the final objective (see Table 2):
  - Binary tournament for *selection*.

Table 2. NSGA-II Parameters

Genetic operator	
Selection	Binary Tournament
Crossover	HUX P_c=0.9
Mutation	Classical P_m=0.7
Other options	
Population size	100
Evaluations	50000

- HUX [41] is used to *crossover* with probability P_c. The HUX crossover exactly interchanges half the alleles that are different in the parents (the genes to be crossed are randomly selected among those that are different in the parents). This operator ensures the maximum distance from the offspring to their parents (exploration).
- Classical *mutation* with probability P_m. This operator changes a gene value at random, sets a gene to zero with probability P_m and sets to one with probability $1 - P_m$. This operator was proposed for rule selection in [42] and it promotes the elimination of the rules, since all individuals of the initial population contained all candidates' rules.

- In addition, if one individual (subset of candidates' rules) does not cover some examples previously covered, then fitness objectives are *penalized*. Then these solutions go (at least) to the second non-dominated front.
- The *stopping criterion* is the number of evaluations.

The *implementation* of the NSGA-II algorithm considered can be reached from Kanpur Genetic Algorithms Laboratory web page [b], adapting some genetic operators and the evaluation of the fitness function. Table 2 shows the parameters used to run NSGA-II.

4.1.3. *Pareto Front Analysis*

The Pareto fronts are generated for each trial and three representative models (according to the objectives accuracy and interpretability) are considered to be analyzed [43,44]:

(1) The most interpretable model: Best $Inter_C$.
(2) The most accurate model: Best Acc.
(3) The median model: Median $Acc - Inter_C$.

[b] http://www.iitk.ac.in/kangal/codes.shtml

5. Experimental Study: Results and Analysis

In order to check the performance of the proposal introduced in this work, nine real-world data sets from the KEEL Project [45,46] [c] have been used:

(1) Plastic Strength (PLA): 3 variables, 1650 records.
(2) Quake (QUA): 4 variables, 2178 records.
(3) Electrical Maintenance (ELE): 5 variables, 1056 records.
(4) Abalone (ABA): 9 variables, 4177 records.
(5) Stock prices (STP): 10 variables, 950 records.
(6) Weather Ankara (WAN): 10 variables, 1609 records.
(7) Weather Izmir (WIZ): 10 variables, 1461 records.
(8) Mortgage (MOR): 16 variables, 1049 records.
(9) Treasury (TRE): 16 variables, 1049 patterns.

First of all, the base fuzzy models are generated by the FasArt neuro-fuzzy algorithm. Next, the multi-objective rule selection is carried out, generating a Pareto Front for each dataset and for each trial, as shown in Algorithm 1: for all the experiments, a fivefold cross validation model is adopted (each fold contained 20% of the records), using four folds for training and one for testing. For each of the possible five different partitions (train/test), both stages of the algorithm were run 6 times, considering a different seed for the random-number generator each time. Therefore, we consider the average results of 30 runs on only three representative models from the Pareto front: Best Interpretability, Best Accuracy and Median Accuracy-Interpretability. Finally, non-parametric statistical tests are run to know the general significance of the results in the context of this manuscript: non-parametric Wilcoxon's signed-rank tests [47,48].

5.1. *FasArt Fuzzy Models*

The fuzzy models were generated by FasArt in fivefold cross validation for each regression problem considered (see data set Table). The FasArt parameters used for all the cases are shown in Table 1. In Table 3, the performance of these fuzzy models is shown: it is possible to see that the accuracy of the models is high (as it is usual for approximative fuzzy modeling approaches). On the other hand, in Tables 4, 5 and 6, the value distribution of relevance rules for each data set (DataSet(I)) and orthogonal transformations can be checked for the three representative models from the Pareto Front. The first line of each table shows the initial/original model (I), while the second line shows the final improved model performance (F). Bold values indicate a lower value in the performance when initial (I) and final (F) models are compared.

In these tables, it is possible to see that the number of rules with "low relevance" selected by the algorithm is higher than expected. Thus, the average values for each

[c] http://sci2s.ugr.es/keel/datasets.php

Table 3. Performance of Fasart Models (according to section 3.1)

Model	Fasart			
	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
PLA	3.483	3.621	0.264	48.6
QUA	0.050	0.054	0.253	119.8
ELE	117867	158820	0.258	92.6
ABA	6.872	7.683	0.265	122.8
STP	2.091	2.270	0.250	101.8
WIZ	5.452	16.555	0.278	221.6
WAN	9.813	21.970	0.273	231.8
MOR	1.041	1.258	0.270	52.6
TRE	0.908	1.339	0.269	49.6

case in the interval [0% − 20%] are: 56.8% (Best $Inter_C$), 58.8% (Median $Acc - Inter_C$) and 62.2% (Best *Acc*) for SVD, 55.9% (Best $Inter_C$), 58.7% (Median $Acc - Inter_C$) and 62.4% (Best *Acc*) for P-QR, and 85.1% (Best $Inter_C$), 85.2% (Median $Acc - Inter_C$) and 87.1% (Best *Acc*) for OLS. This can indicate that rules with lower relevance by orthogonal transformations can be relevant for accuracy goals. On the other hand, to find a value *"gap"* for relevance that allows to discriminate between relevant and not relevant rules in the rule selection can be too complicated if sufficient accuracy for the model is desired.

Table 4. Genetic Rule Influence Distribution by Fasart and SVD Decomposition

	Best $Inter_C$ (%)			Median $Acc - Inter_C$ (%)			Best Acc (%)		
Models	0-20	20-30	30-100	0-20	20-30	30-100	0-20	20-30	30-100
PLA(I)	70.8	13.6	15.6	70.8	13.6	15.6	70.8	13.6	15.6
PLA(F)	34.1	19.8	46.1	47.3	19.9	32.9	62.5	17.3	20.3
QUA(I)	79.0	10.2	10.9	79.0	10.2	10.9	79.0	10.2	10.9
QUA(F)	73.6	12.5	13.9	74.5	12.2	13.3	76.4	11.7	11.9
ELE(I)	97.8	0.2	1.9	97.8	0.2	1.9	97.8	0.2	1.9
ELE(F)	97.6	0.3	2.2	97.7	0.2	2.1	97.7	0.2	2.0
ABA(I)	66.9	14.3	18.7	66.9	14.3	18.7	66.9	14.3	18.7
ABA(F)	62.1	15.5	22.5	61.6	15.4	23.0	65.2	14.1	20.7
STP(I)	28.3	30.1	41.7	28.3	30.1	41.7	28.3	30.1	41.7
STP(F)	24.7	30.7	44.6	24.7	30.3	45.0	26.1	30.5	43.4
WIZ(I)	66.6	17.6	15.8	66.6	17.6	15.8	66.6	17.6	15.8
WIZ(F)	64.2	18.4	17.4	64.5	18.2	17.4	65.0	18.2	16.8
WAN(I)	75.2	14.8	10.0	75.2	14.8	10.0	75.2	14.8	10.0
WAN(F)	72.6	16.2	11.3	72.6	16.1	11.3	73.5	15.7	10.8
MOR(I)	49.1	13.3	37.6	49.1	13.3	37.6	49.1	13.3	37.6
MOR(F)	41.3	15.3	43.4	41.1	15.2	43.7	43.3	15.4	41.3
TRE(I)	51.6	11.3	37.1	51.6	11.3	37.1	51.6	11.3	37.1
TRE(F)	40.8	12.3	46.9	43.4	11.6	45.0	50.4	11.2	38.5

Table 5. Genetic Rule Influence Distribution by Fasart and P-QR Decomposition

	Best $Inter_C$ (%)			Median $Acc - Inter_C$ (%)			Best Acc (%)		
Models	0-20	20-30	30-100	0-20	20-30	30-100	0-20	20-30	30-100
PLA(I)	68.7	14.4	16.9	68.7	14.4	16.9	68.7	14.4	16.9
PLA(F)	35.5	20.4	44.1	48.0	20.0	32.1	61.9	17.7	20.4
QUA(I)	77.5	12.0	10.5	77.5	12.0	10.5	77.5	12.0	10.5
QUA(F)	74.5	12.2	13.3	74.8	11.7	13.5	77.7	10.7	11.7
ELE(I)	97.8	1.1	1.1	97.8	1.1	1.1	97.8	1.1	1.1
ELE(F)	97.6	1.2	1.2	97.6	1.2	1.2	97.7	1.1	1.1
ABA(I)	63.0	16.9	20.0	63.0	16.9	20.0	63.0	16.9	20.0
ABA(F)	55.6	19.3	25.1	59.5	18.0	22.5	62.6	17.2	20.3
STP(I)	28.3	30.5	41.3	28.3	30.5	41.3	28.3	30.5	41.3
STP(F)	25.3	30.4	44.4	25.1	30.5	44.4	27.9	29.4	42.7
WIZ(I)	66.5	17.7	15.8	66.5	17.7	15.8	66.5	17.7	15.8
WIZ(F)	64.0	18.3	17.7	64.2	18.2	17.6	64.9	18.0	17.2
WAN(I)	75.3	14.9	9.8	75.3	14.9	9.8	75.3	14.9	9.8
WAN(F)	72.4	16.1	11.5	72.8	16.0	11.2	73.0	15.9	11.1
MOR(I)	49.1	13.3	37.7	49.1	13.3	37.6	49.1	13.3	37.6
MOR(F)	38.7	13.7	47.6	42.5	13.8	43.6	48.1	13.6	38.3
TRE(I)	51.6	11.7	36.7	51.6	11.7	36.7	51.6	11.7	36.7
TRE(F)	39.3	13.2	47.5	43.5	12.7	43.8	50.9	11.4	37.8

Table 6. Genetic Rule Influence Distribution by Fasart and OLS Decomposition

	Best $Inter_C$ (%)			Median $Acc - Inter_C$ (%)			Best Acc (%)		
Models	0-20	20-30	30-100	0-20	20-30	30-100	0-20	20-30	30-100
PLA(I)	94.2	3.3	2.5	94.2	3.3	2.5	94.2	3.3	2.5
PLA(F)	86.3	7.2	6.5	83.7	8.8	7.5	93.7	2.5	3.8
QUA(I)	96.5	1.5	2.0	96.5	1.5	2.0	96.5	1.5	2.0
QUA(F)	95.0	2.2	2.8	95.0	2.2	2.8	95.5	2.0	2.5
ELE(I)	77.5	14.0	8.4	77.5	14.0	8.4	77.5	14.0	8.4
ELE(F)	75.3	15.4	9.2	75.4	15.4	9.2	76.4	14.8	8.7
ABA(I)	87.5	6.2	6.4	87.5	6.2	6.4	87.5	6.2	6.4
ABA(F)	84.8	7.5	7.7	85.7	6.3	8.0	87.9	5.7	6.5
STP(I)	88.8	5.1	6.1	88.8	5.1	6.1	88.8	5.1	6.1
STP(F)	88.4	5.2	6.5	87.7	5.5	6.8	88.2	5.0	6.8
WIZ(I)	98.6	0.5	0.9	98.6	0.5	0.9	98.6	0.5	0.9
WIZ(F)	98.3	0.6	1.1	98.3	0.6	1.1	98.3	0.6	1.1
WAN(I)	98.4	0.7	1.0	98.4	0.7	1.0	98.4	0.7	1.0
WAN(F)	98.1	0.8	1.1	98.1	0.8	1.1	98.2	0.8	1.0
MOR(I)	76.8	9.1	14.1	76.8	9.1	14.1	76.8	9.1	14.1
MOR(F)	74.9	8.2	16.9	75.5	8.9	15.6	78.3	7.5	14.1
TRE(I)	73.8	8.5	17.7	73.8	8.5	17.7	73.8	8.5	17.8
TRE(F)	65.6	8.1	26.3	66.5	8.7	24.8	72.0	7.5	20.4

5.2. *Genetic Rule Selection: Results*

This section shows the main results obtained by the NSGA-II genetic algorithm and the fitness-functions, that are based on the orthogonal transformations and the accuracy-interpretability trade-off. Tables 7, 8 and 9 show the averaged results obtained from the Pareto Front work over 30 runs for each case study considered: the MSE for training (MSE_{tra}) and testing (MSE_{tst}), the interpretability ($Inter_C$) and the mean rule number (RN). Values in bold indicate a better performance when initial (I) and final (F) models are matched.

In general, these results for the three orthogonal transformations, on three Pareto Front points analyzed (Best InterC, Median Acc-InterC and Best Acc), show that the interpretability have been improved, reducing the complexity and the number of rules of the fuzzy models. On the other hand, the acuracy of the models has been preserved in reasonable levels, without a too much loss of accuracy, and, in some cases, the accuracy has been also improved.

Table 7. Genetic Rule Selection by Fasart and SVD Decomposition

	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
Models	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	3.483	3.621	0.80	48.60
PLA(F)	9.383	9.586	**0.14**	**10.77**	3.820	4.054	**0.24**	**18.77**	**2.620**	**2.779**	**0.46**	**27.43**
Δ	169.50	164.72	-83.03	-77.90	9.73	11.86	-70.27	-61.47	-24.76	-23.34	-42.26	-43.66
QUA(I)	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80
QUA(F)	**0.037**	**0.041**	**0.34**	**80.27**	**0.036**	**0.040**	**0.41**	**85.27**	**0.035**	**0.039**	**0.48**	**88.27**
Δ	-24.91	-24.54	-57.75	-33.01	-27.83	-26.74	-49.16	-28.85	-30.09	-28.31	-39.42	-26.32
ELE(I)	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60	117867	**158820**	**0.76**	92.60
ELE(F)	187127	227572	**0.68**	**81.47**	154594	197875	**0.73**	**84.93**	**116220**	159736	0.76	**88.13**
Δ	63.09	51.72	-10.90	-12.02	27.30	24.06	-4.21	-8.27	-1.42	0.86	0.48	-4.81
ABA(I)	6.872	7.683	0.80	122.80	6.872	7.683	0.80	122.80	6.872	7.683	0.80	122.80
ABA(F)	**6.225**	**7.015**	**0.40**	**93.20**	**5.635**	**6.342**	**0.50**	**95.53**	**5.203**	**5.940**	**0.61**	**102.80**
Δ	-9.01	-8.04	-50.27	-24.14	-17.59	-16.97	-37.58	-22.24	-24.00	-22.49	-23.70	-16.32
STP(I)	**2.091**	**2.270**	0.76	101.80	2.091	**2.270**	0.76	101.80	2.091	2.270	**0.76**	101.80
STP(F)	2.188	2.464	**0.49**	**89.70**	**2.065**	2.373	**0.70**	**88.70**	**1.994**	**2.235**	0.80	**88.13**
Δ	5.99	10.16	-34.93	-11.86	-1.06	5.40	-7.99	-12.84	-4.52	-1.26	5.05	-13.44
WIZ(I)	**5.452**	**16.555**	0.52	221.60	5.452	**16.555**	0.52	221.60	5.452	16.555	0.52	221.60
WIZ(F)	5.593	16.868	**0.44**	**185.67**	**5.292**	16.602	**0.47**	**188.77**	**5.157**	**16.520**	**0.48**	**187.83**
Δ	3.79	1.68	-15.00	-16.30	-2.85	0.19	-9.41	-14.90	-5.84	-0.18	-8.26	-15.33
WAN(I)	9.813	**21.970**	0.44	231.80	9.813	**21.970**	0.44	231.80	9.813	**21.970**	0.44	231.80
WAN(F)	**9.766**	22.206	**0.40**	**189.30**	**9.543**	22.250	**0.41**	**193.93**	**9.343**	22.242	**0.41**	**191.83**
Δ	-0.93	1.36	-10.33	-18.37	-3.10	1.66	-7.84	-16.32	-5.06	1.63	-7.82	-17.22
MOR(I)	1.041	1.258	0.68	52.60	1.041	1.258	0.68	52.60	1.041	1.258	0.68	52.60
MOR(F)	**0.999**	**1.249**	**0.39**	**38.00**	**0.947**	**1.186**	**0.51**	**40.97**	**0.922**	**1.180**	**0.50**	**38.93**
Δ	-2.87	0.44	-42.26	-27.60	-8.30	-5.63	-26.18	-21.79	-10.61	-6.29	-28.83	-25.69
TRE(I)	**0.908**	**1.339**	0.60	49.60	**0.908**	**1.339**	0.60	49.60	0.908	1.339	0.60	49.60
TRE(F)	1.582	1.999	**0.35**	**29.90**	1.016	1.443	**0.50**	**36.87**	**0.884**	**1.326**	**0.58**	**41.27**
Δ	91.81	58.06	-41.57	-39.56	15.27	7.27	-16.29	-25.58	-3.15	-2.02	-3.15	-16.82

Table 8. Genetic Rule Selection by Fasart and P-QR Decomposition

	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
Models	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	3.483	3.621	0.80	48.60
PLA(F)	8.432	8.400	**0.14**	**12.10**	3.911	4.100	**0.24**	**20.37**	**2.620**	**2.746**	**0.48**	**29.20**
Δ	141.85	132.90	-82.78	-75.10	12.31	13.35	-69.81	-58.03	-24.76	-24.17	-39.72	-39.91
QUA(I)	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80
QUA(F)	**0.038**	**0.041**	**0.34**	**78.40**	**0.036**	**0.040**	**0.40**	**81.47**	**0.035**	**0.039**	**0.44**	**87.33**
Δ	-23.19	-23.41	-58.01	-34.65	-27.03	-26.62	-50.35	-32.08	-29.36	-27.87	-45.35	-27.16
ELE(I)	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60	117867	**158820**	**0.76**	92.60
ELE(F)	174284	228080	**0.68**	**81.50**	134434	191295	**0.72**	**84.33**	**116270**	159145	0.77	**87.97**
Δ	53.54	50.24	-10.75	-11.97	15.73	22.68	-4.82	-8.90	-1.38	0.12	0.64	-4.99
ABA(I)	**6.872**	7.683	0.80	122.80	6.872	7.683	0.80	122.80	6.872	7.683	0.80	122.80
ABA(F)	6.928	**7.574**	**0.34**	**83.80**	**5.484**	**6.205**	**0.50**	**97.80**	**5.262**	**5.965**	**0.62**	**103.60**
Δ	1.76	-0.05	-57.65	-31.75	-19.92	-19.04	-37.67	-20.36	-23.08	-22.17	-22.45	-15.68
STP(I)	**2.091**	**2.270**	0.76	101.80	2.091	**2.270**	0.76	101.80	2.091	**2.270**	**0.76**	101.80
STP(F)	2.178	2.427	**0.53**	**87.23**	**2.076**	2.380	**0.75**	**88.57**	**2.001**	2.304	0.81	**87.97**
Δ	3.99	7.48	-30.70	-14.31	-0.67	5.71	-1.80	-12.98	-4.21	2.55	6.41	-13.59
WIZ(I)	**5.452**	**16.555**	0.52	221.60	5.452	16.555	**0.52**	221.60	5.452	16.555	**0.52**	221.60
WIZ(F)	6.018	17.504	**0.39**	**180.77**	**5.286**	**16.477**	0.53	**186.47**	**5.165**	**16.437**	0.52	**187.27**
Δ	15.64	6.16	-23.69	-18.34	-3.08	-0.49	-0.11	-15.81	-5.68	-0.78	-0.60	-15.53
WAN(I)	9.813	**21.970**	0.44	231.80	9.813	**21.970**	0.44	231.80	9.813	**21.970**	0.44	231.80
WAN(F)	**9.513**	22.548	**0.39**	**190.33**	**9.469**	22.421	**0.41**	**197.53**	**9.331**	22.317	**0.40**	**193.87**
Δ	-3.16	2.77	-11.30	-17.90	-3.67	2.17	-7.63	-14.77	-5.10	1.77	-9.17	-16.35
MOR(I)	**1.041**	**1.258**	0.68	52.60	1.041	**1.258**	0.68	52.60	1.041	1.258	0.68	52.60
MOR(F)	1.290	1.485	**0.35**	**35.47**	**0.993**	1.260	**0.52**	**41.40**	**0.922**	**1.179**	**0.49**	**42.07**
Δ	26.44	20.13	-48.59	-32.51	-3.61	-0.32	-24.76	-21.15	-10.61	-6.35	-30.24	-19.75
TRE(I)	**0.908**	**1.339**	0.60	49.60	**0.908**	**1.339**	0.60	49.60	0.908	1.339	0.60	49.60
TRE(F)	1.583	2.157	**0.34**	**28.77**	1.008	1.466	**0.37**	**34.80**	**0.888**	**1.330**	**0.51**	**41.40**
Δ	78.27	66.68	-42.69	-41.87	12.58	9.93	-37.73	-29.72	-2.87	-1.52	-15.65	-16.53

Table 9. Genetic Rule Selection by Fasart and OLS Decomposition

	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
Models	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	3.483	3.621	0.80	48.60
PLA(F)	10.243	10.231	**0.13**	**10.23**	4.551	4.607	**0.22**	**15.57**	**2.592**	**2.726**	**0.53**	**26.93**
Δ	194.21	183.59	-83.30	-79.03	30.92	27.03	-72.56	-68.10	-25.56	-24.74	-34.24	-44.66
QUA(I)	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80	0.050	0.054	0.80	119.80
QUA(F)	**0.038**	**0.041**	**0.35**	**77.33**	**0.036**	**0.039**	**0.42**	**81.90**	**0.035**	**0.038**	**0.46**	**85.70**
Δ	-24.50	-24.49	-56.55	-35.43	-28.26	-27.87	-47.75	-31.63	-30.02	-28.90	-42.50	-28.48
ELE(I)	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60	117867	**158820**	**0.76**	92.60
ELE(F)	153284	212925	**0.69**	**83.80**	130250	179441	**0.74**	**84.57**	**116014**	160133	0.77	**87.70**
Δ	32.20	38.75	-8.94	-9.48	11.81	15.39	-3.35	-8.65	-1.62	1.16	1.01	-5.27
ABA(I)	6.872	7.683	0.80	122.80	6.872	7.683	0.80	122.80	6.872	7.683	0.80	122.80
ABA(F)	**5.757**	**6.404**	**0.36**	**88.00**	**5.387**	**6.103**	**0.48**	**95.00**	**5.148**	**5.870**	**0.56**	**102.20**
Δ	-15.96	-16.38	-54.67	-28.39	-21.42	-20.34	-40.10	-22.69	-24.78	-23.40	-29.68	-16.81
STP(I)	**2.091**	**2.270**	0.76	101.80	2.091	**2.270**	0.76	101.80	2.091	**2.270**	**0.76**	101.80
STP(F)	2.150	2.415	**0.51**	**89.97**	**2.048**	2.346	**0.75**	**89.10**	**1.995**	2.294	0.80	**86.17**
Δ	3.13	7.16	-32.43	-11.64	-1.95	4.10	-0.67	-12.47	-4.48	2.14	4.96	-15.37
WIZ(I)	5.452	**16.555**	0.52	221.60	5.452	**16.555**	0.52	221.60	5.452	16.555	0.52	221.60
WIZ(F)	**5.333**	16.645	**0.45**	**183.03**	**5.251**	16.561	**0.49**	**183.90**	**5.128**	**16.460**	**0.49**	**183.17**
Δ	-1.70	0.80	-12.68	-17.38	-3.67	0.23	-6.02	-17.08	-6.37	-0.53	-6.61	-17.44
WAN(I)	9.813	**21.970**	0.44	231.80	9.813	**21.970**	0.44	231.80	9.813	21.970	0.44	231.80
WAN(F)	**9.648**	22.328	**0.38**	**191.30**	**9.530**	22.132	**0.41**	**199.60**	**9.344**	**21.963**	**0.40**	**195.07**
Δ	-1.94	1.12	-12.43	-17.50	-2.96	0.50	-6.55	-13.90	-5.01	-0.17	-8.84	-15.84
MOR(I)	**1.041**	**1.258**	0.68	52.60	1.041	1.258	0.68	52.60	1.041	1.258	0.68	52.60
MOR(F)	1.097	1.336	**0.36**	**36.83**	**0.965**	**1.221**	**0.52**	**41.13**	**0.921**	**1.174**	**0.50**	**39.83**
Δ	6.89	7.05	-46.03	-29.90	-6.61	-3.16	-24.33	-21.78	-10.73	-6.75	-27.82	-24.00
TRE(I)	**0.908**	**1.339**	0.60	49.60	**0.908**	**1.339**	0.60	49.60	0.908	1.339	0.60	49.60
TRE(F)	1.675	2.076	**0.34**	**26.63**	1.000	1.390	**0.42**	**32.67**	**0.895**	**1.333**	**0.56**	**41.10**
Δ	94.76	51.29	-44.06	-46.18	15.31	4.57	-29.29	-34.03	-2.12	-1.50	-6.97	-17.12

Then, for checking the scope of orthogonal transformations using genetic algorithms, the Wilcoxon test is run on error and interpretability/complexity indices for the three characteristic models from the Pareto front. This test is used for detecting significant differences between two sample means: it is analogous to the paired t-test in non-parametric statistical procedures. In general, the test asks (H_0): do two samples come from populations with the same distributions?. It is based on ranks of the differences between pairs of data.

The Wilcoxon test for the three orthogonal transformations (Tables 10, 11 and 12) accepts that:

- Best $Inter_C$ models have improved the interpretability and complexity index preserving the accuracy of the original models except in QRP.
- Median $Acc - Inter_C$ models have an accuracy similar to the original models, and the interpretability and complexity indexes are improved. So the accuracy has been preserved without relevant loss of precision.
- Best *Acc* models have improved the interpretability and complexity index and the accuracy is preserved.

Table 10. Wilcoxon test for SVD: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	12.0	33.0	Accepted	0.214
NR	45.0	0.0	Rejected	0.008
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	20.0	25.0	Accepted	0.767
NR	45.0	0.0	Rejected	0.008
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	39.0	6.0	Rejected	0.051
NR	45.0	0.0	Rejected	0.008
$Inter_C$	40.0	5.0	Rejected	0.038

In general, for the three orthogonal transformations taken into account, and for the three Pareto Front points analyzed (*Best* $Inter_C$, *Median* $Acc - Inter_C$ and *Best Acc*), the rule selection has generated rule subsets that have improved their interpretability, reducing their complexity and preserving a reasonable level of accuracy in comparison with other works involving the same data sets. In some cases, the accuracy has been improved through the rule selection, simultaneously improving its interpretability. This is more notorious in the Best *Acc* point in comparison with the others: *Best* $Inter_C$ and *Median* $Acc - Inter_C$. In this Best *Acc* point, in averaged values for all data sets, the accuracy ($\Delta MSE_{tra} \cong -9$, $\Delta MSE_{tst} \cong -9$),

Table 11. Wilcoxon test for QRP: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	7.0	38.0	Rejected	0.066
NR	45.0	0.0	Rejected	0.008
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	18.0	27.0	Accepted	0.594
NR	45.0	0.0	Rejected	0.008
$Inter_C$	44.0	1.0	Rejected	0.011
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	35.0	10.0	Accepted	0.139
NR	45.0	0.0	Rejected	0.008
$Inter_C$	38.0	7.0	Rejected	0.066

Table 12. Wilcoxon test for OLS: original model (R+) and improved model (R-)

Best $Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	11.0	34.0	Accepted	0.173
NR	45.0	0.0	Rejected	0.008
$Inter_C$	45.0	0.0	Rejected	0.008
Median $Acc - Inter_C$				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	19.0	26.0	Accepted	0.678
NR	45.0	0.0	Rejected	0.008
$Inter_C$	45.0	0.0	Rejected	0.008
Best Acc				
Measure	R+	R-	Hypothesis (alpha=0.10)	p-value
MSE_{tst}	36.0	9.0	Accepted	0.110
NR	45.0	0.0	Rejected	0.008
$Inter_C$	39.0	6.0	Rejected	0.051

interpretability ($\Delta Inter_C \cong -16$) and the number of rules ($\Delta RN \cong -19$) have been improved for every orthogonal transformation used.

On the other hand, the number of fuzzy rules is decreased from the point of *Best* $Inter_C$ and to the point *Best Acc*, and their accuracy too.

This can be connected with the role of lower influence rules by orthogonal transformations (lower values): in Tables 4, 5 and 6, the distribution of these values is shown by intervals. In general, the genetic rule selection has to choose a relevant number of rules whose importance values are low: according to the usual rule selection by orthogonal transformations, these rules must not be considered. The number of this rule type is increased from the point of *Best* $Inter_C$ to the point *Best Acc*, in proportion to the increase in the accuracy. Thus, these rules have an accuracy role for the fuzzy models.

Another point to be analyzed is the level of influence preserved, i.e., the aggregation of the individual influence of the rules selected in the simplified models: in Tables 13, 14 and 15, the average values are shown for each data set, each Pareto Front Point considered and each orthogonal transformation. In general, this value is around $80\% - 90\%$, there are no relevant differences between the several cases shown, so the selection has saved most of the information of the original model. This can give an idea of how to define a criterion for manual or automatic rule selection.

Table 13. Genetic Influence Preservation Rate by Fasart and SVD Decomposition

Models	Best $Inter_C$ mean(std)	Median $Acc - Inter_C$ mean(std)	Best Acc mean(std)
PLA(F)	0.481(0.075)	0.656(0.079)	0.713(0.070)
QUA(F)	0.799(0.057)	0.824(0.049)	0.799(0.045)
ELE(F)	0.935(0.031)	0.956(0.023)	0.964(0.017)
ABA(F)	0.843(0.051)	0.876(0.036)	0.883(0.028)
STP(F)	0.936(0.041)	0.925(0.036)	0.897(0.039)
WIZ(F)	0.874(0.058)	0.887(0.044)	0.871(0.041)
WAN(F)	0.855(0.068)	0.874(0.055)	0.851(0.051)
MOR(F)	0.815(0.090)	0.880(0.065)	0.803(0.063)
TRE(F)	0.716(0.111)	0.871(0.106)	0.867(0.079)

Table 14. Genetic Influence Preservation Rate by Fasart and P-QR Decomposition

Models	Best $Inter_C$ mean(std)	Median $Acc - Inter_C$ mean(std)	Best Acc mean(std)
PLA(F)	0.524(0.087)	0.679(0.055)	0.721(0.064)
QUA(F)	0.724(0.065)	0.763(0.051)	0.743(0.043)
ELE(F)	0.937(0.024)	0.952(0.017)	0.960(0.014)
ABA(F)	0.773(0.042)	0.853(0.047)	0.846(0.036)
STP(F)	0.907(0.037)	0.924(0.040)	0.889(0.037)
WIZ(F)	0.861(0.050)	0.885(0.036)	0.878(0.026)
WAN(F)	0.869(0.038)	0.896(0.038)	0.874(0.031)
MOR(F)	0.797(0.057)	0.871(0.058)	0.799(0.064)
TRE(F)	0.738(0.082)	0.832(0.049)	0.866(0.091)

At this point, if this selection is carried out by hand, considering these preserved values of influence and using the well-known criterion to select the rules with most influence (value) until this aggregated level is achieved for each data set, then the results are shown in Tables 16, 17 and 18. These tables show the mean of each individual measurement: the MSE for training (MSE_{tra}) and testing (MSE_{tst}), the interpretability ($Inter_C$) and the mean rule number (RN). Values in bold indicate a lower value in the performance when initial (I) and final (F) models are compared.

Table 15. Genetic Influence Preservation Rate by Fasart and OLS Decomposition

Models	Best $Inter_C$ mean(std)	Median $Acc - Inter_C$ mean(std)	Best Acc mean(std)
PLA(F)	0.435(0.181)	0.707(0.078)	0.721(0.048)
QUA(F)	0.791(0.132)	0.865(0.090)	0.812(0.124)
ELE(F)	0.976(0.018)	0.977(0.022)	0.969(0.037)
ABA(F)	0.851(0.036)	0.902(0.029)	0.851(0.049)
STP(F)	0.922(0.046)	0.946(0.038)	0.903(0.040)
WIZ(F)	0.902(0.043)	0.917(0.035)	0.886(0.024)
WAN(F)	0.902(0.058)	0.928(0.038)	0.889(0.060)
MOR(F)	0.781(0.099)	0.829(0.086)	0.736(0.063)
TRE(F)	0.736(0.108)	0.845(0.074)	0.916(0.043)

Here, it is possible to check that the accuracy is worse than the selection carried out by the genetic algorithm, also the number of rules is lower. This is connected with the role of the rule associated with lower values of influence by orthogonal transformations. This can be due to the conservative behavior of the genetic selection, perhaps, in some of the cases involves in this work, other more risky selections could be possible, reducing the number of rules, on the basis of lower influential rules and reducing the accuracy, but keeping it within competitive levels.

Table 16. Rule Selection based on Aggregated Influence Rule Level (β) and SVD

Models	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
	$\beta = 48.1$				$\beta = 65.6$				$\beta = 71.3$			
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60
PLA(F)	12.903	13.217	**0.34**	**5.40**	11.501	11.707	**0.42**	**9.80**	10.268	10.409	**0.35**	**11.80**
Δ	270.43	265.05	-58.12	-88.89	230.16	223.35	-47.47	-79.84	194.77	187.51	-55.93	-75.72
	$\beta = 79.9$				$\beta = 82.4$				$\beta = 79.9$			
QUA(I)	**0.050**	**0.054**	0.80	119.80	**0.050**	**0.054**	0.80	119.80	**0.050**	**0.054**	0.80	119.80
QUA(F)	0.082	0.081	**0.76**	**46.40**	0.084	0.083	**0.73**	**50.40**	0.082	0.081	**0.76**	**46.40**
Δ	63.33	49.71	-5.45	-61.27	68.74	54.30	-8.89	-57.93	63.33	49.71	-5.45	-61.27
	$\beta = 93.5$				$\beta = 95.6$				$\beta = 96.4$			
ELE(I)	**117867**	**158820**	**0.76**	92.60	**117867**	**158820**	**0.76**	92.60	**117867**	**158820**	**0.76**	92.60
ELE(F)	805024	896402	0.78	**67.20**	710215	842158	0.79	**72.20**	649513	794805	0.79	**74.40**
Δ	583.00	464.42	2.91	-27.43	502.56	430.26	3.99	-22.03	451.06	400.45	3.68	-19.65
	$\beta = 84.3$				$\beta = 87.6$				$\beta = 88.3$			
ABA(I)	**6.872**	**7.683**	0.80	122.80	**6.872**	**7.683**	0.80	122.80	**6.872**	**7.683**	0.80	122.80
ABA(F)	14.686	14.849	**0.69**	**67.20**	13.617	13.856	**0.66**	**75.80**	11.258	11.696	**0.68**	**77.40**
Δ	113.72	93.27	-13.78	-45.28	98.16	80.35	-17.08	-38.27	63.82	52.24	-15.59	-36.97
	$\beta = 93.6$				$\beta = 92.5$				$\beta = 89.7$			
STP(I)	**2.091**	**2.270**	0.76	101.80	**2.091**	**2.270**	0.76	101.80	**2.091**	**2.270**	0.76	101.80
STP(F)	6.222	5.647	**0.73**	**65.80**	6.675	6.206	**0.75**	**64.00**	6.786	6.280	**0.68**	**60.80**
Δ	197.59	148.82	-4.29	-35.36	219.24	173.45	-1.31	-37.13	224.55	176.72	-10.96	-40.28
	$\beta = 87.4$				$\beta = 88.7$				$\beta = 87.1$			
WIZ(I)	**5.452**	**16.555**	0.52	221.60	**5.452**	**16.555**	0.52	221.60	**5.452**	**16.555**	0.52	221.60
WIZ(F)	15.944	27.400	**0.36**	**159.40**	14.003	24.347	**0.36**	**163.20**	16.539	27.968	**0.36**	**158.40**
Δ	192.44	65.51	-30.97	-28.07	156.84	47.07	-30.42	-26.35	203.34	68.94	-31.07	-28.52
	$\beta = 85.5$				$\beta = 87.4$				$\beta = 85.1$			
WAN(I)	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80
WAN(F)	36.867	52.509	**0.34**	**160.80**	35.373	51.420	**0.34**	**166.60**	37.111	52.614	**0.34**	**159.40**
Δ	275.70	139.01	-23.36	-30.63	260.47	134.05	-22.08	-28.13	278.19	139.49	-23.62	-31.23
	$\beta = 81.5$				$\beta = 88.0$				$\beta = 80.3$			
MOR(I)	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60
MOR(F)	3.471	3.600	0.81	**23.20**	2.773	2.756	0.83	**28.40**	3.488	3.643	0.85	**22.20**
Δ	233.26	186.17	18.59	-55.89	166.29	119.10	21.51	-46.01	234.92	189.61	25.65	-57.79
	$\beta = 71.6$				$\beta = 87.1$				$\beta = 86.7$			
TRE(I)	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60
TRE(F)	6.183	6.257	2.78	**15.80**	4.326	3.929	2.58	**24.60**	4.326	3.929	2.58	**24.40**
Δ	580.70	367.21	363.95	-68.15	376.27	193.36	330.72	-50.40	376.27	193.36	330.55	-50.81

Table 17. Rule Selection based on Aggregated Influence Rule Level (β) and P-QR

	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
Models	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
		$\beta = 52.4$				$\beta = 67.9$				$\beta = 72.1$		
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60
PLA(F)	10.542	10.591	**0.34**	**7.80**	10.360	10.421	**0.75**	**12.80**	10.713	10.798	**0.68**	**14.40**
Δ	202.63	192.54	-57.46	-83.95	197.40	187.83	-5.77	-73.66	207.56	198.25	-14.48	-70.37
		$\beta = 72.4$				$\beta = 76.3$				$\beta = 74.3$		
QUA(I)	**0.050**	0.054	**0.80**	119.80	**0.050**	0.054	0.80	119.80	**0.050**	0.054	**0.80**	119.80
QUA(F)	0.052	**0.054**	0.90	**39.20**	0.052	**0.054**	**0.79**	**44.40**	0.052	**0.054**	0.84	**41.60**
Δ	4.33	-0.61	12.60	-67.28	3.98	-0.44	-1.14	-62.94	4.13	-0.63	5.50	-65.28
		$\beta = 93.7$				$\beta = 95.2$				$\beta = 96.0$		
ELE(I)	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60
ELE(F)	435932	585178	**0.60**	**70.20**	373566	455260	**0.62**	**73.60**	341114	410057	**0.64**	**75.80**
Δ	269.85	268.45	-21.00	-24.19	216.94	186.65	-17.83	-20.52	189.41	158.19	-15.66	-18.14
		$\beta = 77.3$				$\beta = 85.3$				$\beta = 84.6$		
ABA(I)	**6.872**	**7.683**	**0.80**	122.80	**6.872**	**7.683**	**0.80**	122.80	**6.872**	**7.683**	0.80	122.80
ABA(F)	8.494	8.822	0.97	**57.00**	8.106	8.508	0.83	**73.40**	8.049	8.434	**0.79**	**71.80**
Δ	23.61	14.82	21.76	-53.58	17.95	10.74	4.05	-40.23	17.13	9.77	-0.94	-41.53
		$\beta = 90.7$				$\beta = 92.4$				$\beta = 88.9$		
STP(I)	**2.091**	**2.270**	**0.76**	101.80	**2.091**	**2.270**	**0.76**	101.80	**2.091**	**2.270**	**0.76**	101.80
STP(F)	6.129	6.493	0.98	**63.00**	5.830	6.133	0.94	**65.00**	6.491	6.751	1.02	**60.80**
Δ	193.14	186.08	28.45	-38.11	178.83	170.22	23.53	-36.15	210.44	197.47	34.73	-40.28
		$\beta = 86.1$				$\beta = 88.5$				$\beta = 87.8$		
WIZ(I)	**5.452**	**16.555**	0.52	221.60	**5.452**	**16.555**	0.52	221.60	**5.452**	**16.555**	0.52	221.60
WIZ(F)	16.773	28.659	**0.41**	**156.80**	16.225	28.503	**0.41**	**163.80**	16.334	28.507	**0.41**	**161.80**
Δ	207.64	73.11	-21.88	-29.24	197.59	72.17	-21.77	-26.08	199.58	72.20	-21.77	-26.99
		$\beta = 86.9$				$\beta = 89.6$				$\beta = 87.4$		
WAN(I)	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80
WAN(F)	18.654	30.660	**0.38**	**166.20**	17.844	30.445	**0.38**	**174.80**	18.450	30.656	**0.38**	**167.80**
Δ	90.09	39.56	-13.39	-28.30	81.84	38.58	-12.57	-24.59	88.01	39.54	-13.29	-27.61
		$\beta = 79.7$				$\beta = 87.1$				$\beta = 79.9$		
MOR(I)	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60
MOR(F)	3.734	3.962	1.30	**23.00**	2.417	2.462	0.97	**28.40**	3.734	3.962	1.30	**23.00**
Δ	258.59	214.92	90.60	-56.27	132.11	95.71	43.21	-46.01	258.59	214.92	90.60	-56.27
		$\beta = 73.8$				$\beta = 83.2$				$\beta = 86.6$		
TRE(I)	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60
TRE(F)	6.228	6.535	0.95	**17.80**	4.983	5.133	1.13	**23.00**	4.683	4.632	0.91	**25.20**
Δ	585.62	387.99	58.10	-64.11	448.60	283.29	89.04	-53.63	415.57	245.85	51.83	-49.19

Table 18. Rule Selection based on Aggregated Influence Rule Level (β) and OLS

	Best $Inter_C$				Median $Acc - Inter_C$				Best Acc			
Models	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN	MSE_{tra}	MSE_{tst}	$Inter_C$	RN
		$\beta = 43.5$				$\beta = 70.7$				$\beta = 72.1$		
PLA(I)	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60	**3.483**	**3.621**	0.80	48.60
PLA(F)	12.834	12.912	**0.27**	**2.40**	15.459	15.648	**0.34**	**7.80**	15.415	15.607	**0.34**	**8.00**
Δ	268.42	256.63	-66.36	-95.06	343.78	332.20	-57.40	-83.95	342.52	331.08	-57.30	-83.54
		$\beta = 79.1$				$\beta = 86.5$				$\beta = 81.2$		
QUA(I)	0.050	0.054	**0.80**	119.80	0.050	0.054	**0.80**	119.80	0.050	0.054	**0.80**	119.80
QUA(F)	**0.044**	**0.045**	0.92	**17.60**	**0.046**	**0.048**	0.91	**25.40**	**0.044**	**0.046**	0.90	**19.60**
Δ	-12.39	-17.38	14.83	-85.31	-7.64	-10.96	13.94	-78.80	-10.96	-15.07	12.53	-83.64
		$\beta = 97.6$				$\beta = 97.7$				$\beta = 96.9$		
ELE(I)	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60	**117867**	**158820**	0.76	92.60
ELE(F)	421006	606827	**0.61**	**65.60**	416034	599132	**0.61**	**66.40**	464177	637139	**0.57**	**62.40**
Δ	257.19	282.09	-19.49	-29.16	252.97	277.24	-19.50	-28.29	293.82	301.17	-24.37	-32.61
		$\beta = 85.1$				$\beta = 90.2$				$\beta = 85.1$		
ABA(I)	**6.872**	7.683	**0.80**	122.80	6.872	7.683	**0.80**	122.80	**6.872**	7.683	**0.80**	122.80
ABA(F)	6.934	**7.120**	1.01	**31.40**	**6.702**	**6.936**	1.06	**39.80**	6.934	**7.120**	1.01	**31.40**
Δ	0.91	-7.33	25.79	-74.43	-2.47	-9.73	32.80	-67.59	0.91	-7.33	25.79	-74.43
		$\beta = 92.2$				$\beta = 94.6$				$\beta = 90.3$		
STP(I)	**2.091**	**2.270**	**0.76**	101.80	**2.091**	**2.270**	**0.76**	101.80	**2.091**	**2.270**	**0.76**	101.80
STP(F)	21.909	22.157	1.75	**42.00**	20.390	20.585	1.56	**47.20**	22.529	22.724	2.01	**38.60**
Δ	947.83	876.22	129.80	-58.74	875.16	806.97	105.46	-53.63	977.48	901.23	164.36	-62.08
		$\beta = 90.2$				$\beta = 91.7$				$\beta = 88.6$		
WIZ(I)	**5.452**	**16.555**	**0.52**	221.60	**5.452**	**16.555**	0.52	221.60	**5.452**	**16.555**	**0.52**	221.60
WIZ(F)	28.209	33.503	0.55	**92.40**	26.031	31.157	**0.51**	**100.00**	28.875	33.876	0.59	**85.20**
Δ	417.40	102.37	5.51	-58.30	377.45	88.20	-1.51	-54.87	429.60	104.63	13.84	-61.55
		$\beta = 90.2$				$\beta = 92.8$				$\beta = 88.9$		
WAN(I)	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80	**9.813**	**21.970**	0.44	231.80
WAN(F)	30.548	32.645	**0.30**	**102.80**	29.293	32.079	**0.39**	**119.00**	30.985	33.248	**0.30**	**96.20**
Δ	211.31	48.59	-30.99	-55.65	198.52	46.02	-11.71	-48.66	215.76	51.34	-32.21	-58.50
		$\beta = 78.1$				$\beta = 82.9$				$\beta = 73.6$		
MOR(I)	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60	**1.041**	**1.258**	**0.68**	52.60
MOR(F)	5.679	5.741	2.29	**16.20**	5.327	5.453	1.74	**19.00**	6.241	6.531	3.06	**13.80**
Δ	445.33	356.34	236.43	-69.20	411.56	333.44	156.38	-63.88	499.28	419.15	350.51	-73.76
		$\beta = 73.6$				$\beta = 84.5$				$\beta = 91.6$		
TRE(I)	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60	**0.908**	**1.339**	**0.60**	49.60
TRE(F)	7.148	7.177	3.52	**11.00**	6.329	6.341	3.47	**16.00**	5.880	5.969	3.20	**21.40**
Δ	686.97	435.91	485.86	-77.82	596.81	373.47	478.37	-67.74	547.38	345.69	433.40	-56.85

The fact that there are no significant differences in the results among the orthogonal transformations, despite the different nature of each one, can be due to the trade-off considered through the genetic rule selection. The consideration of this trade-off can compensate the different attention paid by each orthogonal transformation to rule antecedents or outputs.

6. Conclusions

This work is focused on the checking of the capacities and drawbacks of orthogonal transformations for complexity reduction and interpretability improving of fuzzy models. This aim is carried out by rule selection using a genetic algorithm subject to accuracy-interpretability trade-off, and the influence rule provided by orthogonal transformations. Three of these transformations have been used: SVD, P-QR and OLS, each of which has its own behavior in this task.

In order to check this, nine regression problems have been involved in the experimental work. The results achieved by the genetic selection on complexity, interpretability and accuracy are reasonable, but a bit conservative from the number of rules-accuracy point of view. The experiments have demonstrated the relevance of the rules associated with lower influence values by orthogonal transformations, so these rules will not be considered in the traditional management of rule selection, but they have relevance from the accuracy point of view. In the Best *Acc* Pareto Front point it has been possible, in averaged values for all data sets, to improve simultaneously: the training error (MSE_{tra}), the test error (MSE_{tst}), the interpretability ($Inter_C$) and the number of rules (RN).

On the other hand, some references on the level rule influence to be preserved in the simplified fuzzy model have been obtained. This can be used in future for the definition of the criteria to select the fuzzy rules.

Another aspect to be remarked on is that there are no significant differences between the orthogonal transformations considered in this work, despite their different natures and behavior. This can be explained by the consideration of the accuracy-interpretability trade-off through the genetic rule selection.

Acknowledgment

This work was supported by the Spanish Ministry of Science and Innovation under grants no. DPI2009-14410-C02-02 and IPT-2011-1656-370000.

References

1. H. Wu and J. Mendel, "Classification of battlefield ground vehicles using acoustic features and fuzzy logic rule-based classifiers," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 56 – 72, February 2007.
2. J. Garibaldi and T. Ozen, "Uncertain fuzzy reasoning: A case study in modelling expert decision making," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 16 – 30, February 2007.

3. R. John and P. Innocent, “Modeling uncertainty in clinical diagnosis using fuzzy logic,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 35, no. 6, pp. 1340 – 1350, December 2005.
4. A. Konar, *Computational Intelligence: Principles, techniques and applications.* Springer-Verlag, Berlin, 2005.
5. F. Karray and C. d. De Silva, *Soft Computing and Intelligent Systems Design. Theory, Tools and Applications.* Addison Wesley, 2004.
6. M. Gacto, R. Alcalá, and F. Herrera, “Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures,” *Information Sciences*, vol. 181, pp. 4340 – 4360, 2011.
7. H. Ishibuchi and Y. Nojima, “Discussions on interpretability of fuzzy systems using simple examples,” *Proc. of 13th IFSA World Congress and 6th Conference of EUSFLAT*, pp. 1649–1654, 2009.
8. J. Alonso, L. Magdalena, and G. González-Rodríguez, “Looking for a good fuzzy system interpretability index: An experimental approach,” *International Journal of Approximate Reasoning*, vol. 51, no. 1, pp. 115 – 134, December 2009.
9. S.-M. Zhou, J. M. Garibaldi, R. I. John, and F. Chiclana, “On constructing parsimonious type-2 fuzzy logic systems via influential rule selection,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 3, pp. 654 – 667, June 2009.
10. C. Mencar and A. Fanelli, “Interpretability constraints for fuzzy information granulation,” *Information Sciences*, vol. 178, no. 24, pp. 4585 – 4618, December 2008.
11. G. H. Golub and C. F. Van Loan, *Matrix computations.* The Johns Hopkins University Press, 1996 (Third edition).
12. A. Laub, “Numerical linear algebra aspects of control design computations,” *IEEE Transactions Automat. Contr.*, vol. 30, pp. 97–108, 1985.
13. S.-M. Zhou and J. Q. Gan, “Constructing l2-svm-based fuzzy classifiers in high-dimensional space with automatic model selection and fuzzy rule ranking,” *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 3, pp. 398 – 409, June 2007.
14. M. Setnes and R. Babuška, “Rule base reduction: Some comments on the use of orthogonal transforms,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 31, no. 2, pp. 199 – 206, May 2001.
15. H. Roubos and M. Setnes, “Compact and transparent fuzzy models and classfiers through iterative complexity reduction,” *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 516–524, August 2001.
16. J. Yen and L. Wang, “Simplifying fuzzy rule-based models using orthogonal transformation methods,” *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, vol. 29, no. 1, pp. 13–24, February 1999.
17. G. Mouzouris and M. Mendel, “Designing fuzzy logic systems for uncertain environments using a singular-value-qr decomposition,” in *The 5th IEEE International Conference on Fuzzy Systems. FUZZ IEEE 1996*, New Orleans, LA, September 1996, pp. 295 – 301.
18. L. Wang and R. Langari, “Building sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques,” *IEEE Transactions on Fuzzy Systems*, vol. 3, no. 4, pp. 454–458, November 1995.
19. S. Van Huffel and J. Vandewalle, “Subset selection using the total least squares approach in collinearity problems with errors in the variables,” *Linear Algebra and its applications*, vol. 88/89, pp. 695–714, 1987.
20. S.-M. Zhou and J. Q. Gan, “Low-level interpretability and high-level interpretability: a unified view of data-driven interpretable fuzzy system modelling,” *Fuzzy Sets and Systems*, vol. 159, pp. 3091 – 3131, 2008.

21. F. Herrera, "Genetic fuzzy systems: Taxonomy, current research trends and prospects." *Evolutionary Intelligence*, vol. 1, pp. 27 – 46, 2008.
22. R. Alcalá, J. Alcalá-Fdez, J. Casillas, O. Cordón, and F. Herrera, "Hybrid learning models to get the interpretability-accuracy trade-off in fuzzy modeling," *Soft Computing*, vol. 10, no. 9, pp. 717 – 734, 2006.
23. J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Accuracy Improvements in Linguistic Fuzzy Modelling*, ser. Studies in Fuzziness and SoftComputing. Berlin Heildelberg: Springer-Verlag, 2003, vol. 129, ch. Accuracy Improvements to Find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview, pp. 3–24.
24. H. Ishibuchi, Y. Kaisho, and Y. Nojima, "Complexity, interpretability and explanation capability of fuzzy rule-based classifiers," in *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*, 20-24 August 2009, pp. 1730 – 1735.
25. Y. Jin, "Fuzzy Modeling of High-Dimensional Systems: Complexity Reduction and Interpretability Improvement," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 2, pp. 212–221, April 2000.
26. M. Setnes, R. Babuška, U. Kaymak, and H. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Transactions on Systems, Man and Cybernetics. Part B: Cybernetics*, vol. 28, no. 3, pp. 376 – 386, June 1998.
27. H. Ishibuchi, Y. Nakashima, and Y. Nojima, "Search ability of evolutionary multiobjective optimization algorithms for multiobjective fuzzy genetics-based machine learning," in *IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE 2009*, 20-24 August 2009, pp. 1724 – 1729.
28. A. Lotfi, M. Howarth, and J. Hull, "Orthogonal fuzzy rule-based systems: Selection of optimum rules," *Neural Computing & Applications*, vol. 9, no. 1, pp. 4–11, 2000.
29. J. Cano Izquierdo, Y. Dimitriadis, E. Gómez Sánchez, and J. López Coronado, "Learnning from noisy information in FasArt and Fasback neuro-fuzzy systems," *Neural Networks*, vol. 14, no. 4-5, pp. 407–425, May 2001.
30. G. Sainz Palmero, Y. Dimitriadis, J. Cano Izquierdo, E. Gómez Sánchez, and E. Parrado Hernández, "ART based model set for pattern recognition: FasArt family," in *Neuro-fuzzy pattern recognition*, H. Bunke and A. Kandel, Eds. World Scientific Pub. Co., December 2000, ch. 1, pp. 147–177.
31. M. Galende-Hernández, G. Sainz-Palmero, and M. Fuente-Aparicio, "Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection," *Soft Computing*, vol. 16, Issue 3, pp. 451–470, 2012.
32. Y. Nojima, R. Alcalá, H. Ishibuchi, and F. Herrera, "Special issue on evolutionary fuzzy systems," *Soft Computing*, vol. 15, no. 12, pp. 2299–2301, 2011.
33. R. Alcalá, Y. Nojima, H. Ishibuchi, and F. Herrera, "Special issue on evolutionary fuzzy systems," *International Journal of Computational Intelligence Systems*, vol. 5, no. 2, pp. 209–211, 2012.
34. K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transacions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, April 2002.
35. R. Alcalá, M. Gacto, and F. Herrera, "A fast and scalable multi-objective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems," *IEEE Transactions on Fuzzy Systems*, vol. 19:4, pp. 666 – 681, 2011.
36. J. Casillas, O. Cordón, F. Herrera, and L. Magdalena, *Interpretability Issues in Fuzzy Modelling*, ser. Studies in Fuzziness and SoftComputing. Berlin Heildelberg: Springer-Verlag, 2003, vol. 128, ch. Interpretability Improvements to Find the Balance Interpretability-Accuracy in Fuzzy Modeling: An Overview, pp. 3–22.

37. G. Sainz Palmero, J. Juez Santamaria, E. Moya de la Torre, and J. Perán González, "Fault detection and fuzzy rule extraction in AC motors by a neuro-fuzzy ART-based system," *Engineering Applications of Artificial Intelligence*, vol. 18, pp. 867–874, 2005.
38. G. Sainz, M. Fuente, and P. Vega, "Recurrent neuro-fuzzy modelling of a wastewater treatment plant," *European Journal of Control*, vol. 10, pp. 83–95, 2004.
39. E. Gómez-Sánchez, Y. Dimitriadis, J. Cano-Izquierdo, and J. López-Coronado, "μARTMAP: use of mutual information for category reduction in Fuzzy ARTMAP," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 58 – 69, January 2002.
40. E. Parrado-Hernández, E. Gómez-Sánchez, and Y. Dimitriadis, "Study of distributed learning as a solution to category proliferation in fuzzy artmap based neural systems," *Neural Networks*, vol. 16, no. 7, pp. 1039 – 1057, September 2003.
41. E. L. J., "The chc adaptive search algorithm : How to have safe search when engaging in nontraditional genetic recombination," *Foundations of Genetic Algorithms*, pp. 265–283, 1991.
42. H. Ishibuchi, T. Murata, and I. B. Türksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets and Systems*, vol. 89, no. 2, pp. 135 – 150, July 1997.
43. M. Gacto, R. Alcalá, and F. Herrera, "Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 515 – 531, 2010.
44. R. Alcalá, P. Ducange, F. Herrera, B. Lazzerini, and F. Marcelloni, "A multiobjective evolutionary approach to concurrently learn rule and data bases of linguistic fuzzy-rule-based systems," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 5, pp. 1106 – 1122, October 2009.
45. J. Alcalá-Fdez, L. Sánchez, S. García, M. J. del Jesus, S. Ventura, J. M. Garrell, J. Otero, C. Romero, J. Bacardit, V. M. Rivas, J. C. Fernndez, and F. Herrera, "KEEL: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing - A Fusion of Foundations, Methodologies and Applications*, vol. 13, no. 3, pp. 307 – 318, Febraury 2009.
46. J. Alcalá-Fdez, A. Fernandez, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera, "KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework," *Journal of Multiple-Valued Logic and Soft Computing*, vol. 17:2-3, pp. 255–287, 2011.
47. J. Zar, *Biostatistical Analysis*. Prentice Hall, 1999.
48. D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. Chapman & Hall/CRC, 2003.