Soft Computing (Special issue on Genetic Fuzzy Systems) manuscript No.
(will be inserted by the editor)

Adaptation and Application of Multi-Objective Evolutionary Algorithms for Rule Reduction and Parameter Tuning of Fuzzy Rule-Based Systems ⋆

María José Gacto, Rafael Alcalá, Francisco Herrera ⋆⋆

Dept. Computer Science and A.I., University of Granada, Granada, E-18071, Spain

Received: date / Revised version: date

Abstract Recently, Multi-Objective Evolutionary Algorithms have been applied to improve the difficult trade-off between interpretability and accuracy of Fuzzy Rule-Based Systems. It is known that both requirements are usually contradictory, however, these kinds of algorithms can obtain a set of solutions with different trade-offs.

This contribution analyzes different application alternatives in order to attain the desired accuracy/interpretability balance by maintaining the improved accuracy that a tuning of membership functions could give but trying to obtain more compact models. In this way, we propose the use of Multi-Objective Evolutionary Algorithms as a tool to get almost one improved solution with respect to a classic single objective approach (a solution that could dominate the one obtained by such algorithm in terms of the system error and number of rules). To do that, this work presents and analyzes the application of six different Multi-Objective Evolutionary Algorithms to obtain simpler and still accurate linguistic fuzzy models by performing rule selection and a tuning of the membership functions. The results on two different scenarios show that the use of expert knowledge in the algorithm design process significantly improves the search ability of these algorithms and that they are able to improve both objectives together, obtaining more accurate and at the same time simpler models with respect to the single objective based approach.

1 Introduction

Many automatic techniques have been proposed in the literature to extract a proper set of fuzzy rules from numerical data. Most of these techniques usually try to improve the performance associated to the prediction error without pay a special attention to the system interpretability, an essential aspect of Fuzzy Rule-Based Systems (FRBSs). In the last years, the problem of finding the right trade-off between interpretability and accuracy, in spite of the original nature of fuzzy logic, has arisen a growing interest in methods which take both aspects into account [7,8]. Of course, the ideal thing would be to satisfy both criteria to a high degree, but since they are contradictory issues generally it is not possible. A way to do that, is to improve the system accuracy but trying to maintain the interpretability to an acceptable level [8].

A widely-used technique to improve the accuracy of linguistic Fuzzy Rule-Based Systems (FRBSs) is the *tuning* of Membership Functions (MFs) [1,3,4,8,9], which refines a previous definition of the data base once the rule base has been obtained. The classic approach to perform tuning consists of refining the three definition parameters that identify triangular MFs associated to the labels comprising the initial data base. Although tuning is one of the most powerful techniques to improve the system performance [8,9], sometimes an excessive number of rules is initially considered to reach the highest degree of accuracy. In order to maintain the interpretability to an acceptable level, a recent work [9] has considered the selection of rules together with the tuning of MFs within the same process (not in different stages) and considering performance criteria. In this way, rules are extracted only if it is possible to maintain or even improve the system accuracy. A very interesting conclusion from [9] is that both techniques can present a positive synergy in most of the cases (similar or more accurate models could be obtained by reducing the number of rules) when they are combined within the same process.

On the other hand, since this approach presents a multi-objective nature the use of Multi-Objective Evolutionary Algorithms (MOEAs)[12,17] to obtain a set of

⋆ Supported in part by the Spanish Ministry of Education and Science under grant no. TIN2005-08386-C05-01, and the Andalusian government under grant no. P05-TIC-00531.

⋆⋆ *Present address:* mjgacto@ugr.es and {alcala, herrera}@decsai.ugr.es

solutions with different degrees of accuracy and number of rules could represent an interesting way to work (by considering both characteristics as objectives). Although there are some works in the literature using standard or specific MOEAs to improve the difficult trade-off between interpretability and accuracy of Mamdani FRBSs [10, 14, 26–29, 33], practically all these works were applied to classification problems trying to obtain the complete Pareto (set of non-dominated solutions with different trade-off) by selecting or learning the set of rules better representing the example data, i.e., improving the system classification ability and decreasing the system complexity but not considering learning [2] or tuning [1, 3, 4, 7, 9] of the fuzzy system parameters, which involves another degree of trade-off and type of Pareto front, a more complicated search space and therefore needs different considerations with respect to the works in the existing literature.

Indeed, to directly apply the most recognized MOEAs for general use in order to perform together tuning and rule selection could present some important problems. As David Goldberg stated in [23], the integration of single methods into hybrid intelligent systems goes beyond simple combinations. For him, the future of Computational Intelligence "*lies in the careful integration of the best constituent technologies*", and subtle integration of the abstraction power of fuzzy systems and the innovative power of genetic systems requires a design sophistication that goes further than putting everything together. This is the case when parameter tuning and rule selection are performed by directly applying the most known MOEAs for general use, where several problems arise due to the complex search space concerning this problem.

The main problem is that it is practically impossible to obtain the complete optimal Pareto. This is due to several reasons:

1. There exist a lot of different subsets of rules with more or less the same number of rules (different rule configurations) but representing really different or alternative tuning possibilities.
2. It is easier to decrease the number of rules than to reduce the system error (which is more dependent of the tuning task). This provokes a faster tuning of the simplest solutions before exploring more promising rule configurations (which are dominated by such premature solutions).
3. The obtained parameters (in general) tends to be optimal for these premature solutions making difficult the appearance of better alternative solutions.

In this way, it is necessary to include any expert knowledge in the MOEA design process. An adequate application of standard MOEAs could partially deal with this problem by focusing the search in the most interesting zone of the Pareto frontier. Taking into account that non-dominated solutions with a small number of rules and high errors are not interesting since they have not the desired trade-off between accuracy and interpretability, we could focus the search only in the Pareto zone with the most accurate solutions trying to obtain the least possible number of rules. The best way to do this is to start with solutions selecting all the possible rules, which favours a progressive extraction of bad rules (those that do not improve with the tuning of parameters), only by means of the mutation at the beginning and then by means of the crossover.

A secondary problem is that it is difficult to obtain very accurate solutions by favoring the crossing of solutions with very different rule configurations (those in the Pareto), which should obtain the best accuracy by learning different parameters for the MFs. Although this is not the major problem (MOEAs can obtain good results considering the established mechanisms), significant improvements could be achieved by addressing this problem in the proper way, i.e., standard algorithms can be specifically improved to perform rule selection and tuning together. A way to do that is to establish mating restrictions. However, again it should be done based on the experience by taking into account that exploration and exploitation are both mainly needed at different stages. In this way, a new method was recently proposed in [5], which by modifying the Strength Pareto Evolutionary Algorithm 2 (SPEA2) [44] progressively concentrates the search in the most promising solutions, allowing exploration at first and favoring the exploitation of the most accurate solutions at the end (the Accuracy-Oriented SPEA2, SPEA2$_{Acc}$). Another possibility could be the application of two versions of the well-known Non-dominated Sorting Genetic Algorithm II (NSGA-II) [18] for finding knees [6] (theoretically the most promising Pareto zones in these kinds of problems), since the modifications proposed in [5] were not successful by considering the NSGA-II approach.

Our main aim in this contribution is to analyze different alternatives in order to attain the desired accuracy/interpretability balance by maintaining the improved accuracy that a tuning of MFs could give but trying to obtain more compact models. In this way, we propose the use of MOEAs as the tool to get almost one improved solution with respect to the classic single objective algorithm (a solution that could dominate the one obtained by such algorithm in terms of the system error and the number of rules). To do that, this work presents and analyzes the application of six different MOEAs to obtain simpler and still accurate linguistic fuzzy models by performing rule selection and a classic tuning of the MF parameters (an analysis on the classic tuning could help to extend the better approaches in order to consider other kinds of techniques or new interpretability measures for further works, e.g., another tuning types, learning, etc.). These methods are the well-known SPEA2, NSGA-II and two versions of NSGA-II for finding knees (standard MOEAs adapting and applying proper genetic operators), and two extended MOEAs for specific application, SPEA2$_{Acc}$ and an extension of it proposed in this paper that by applying a more intelligent crossover operator (specific for this problem) is able to extract more useful information from parents with different configurations, SPEA2$_{Acc^2}$. The results on two different scenarios show that the use of expert knowledge in the MOEAs design process significantly improves the search ability of these algorithms.

In order to do that, this contribution is arranged as follows. Next section presents a brief study of the existing MOEAs for general purpose which usually are modified or directly applied to obtain FRBSs with good interpretability-accuracy trade-off. In order to show the main differences with the previous works, Section 3 briefly analyzes the state of the art on the use of MOEAs to get the desired trade-off in different application areas of FRBSs. In Section 4, we describe the different MOEAs and appropriate genetic operators for their proper application. Section 5 shows an experimental study of these methods in two complex and interesting problems. Finally, Section 6 points out some conclusions and further research lines.

2 Multi-Objective Evolutionary Algorithms

Evolutionary algorithms simultaneously deal with a set of possible solutions (the so-called population) which allows to find several members of the Pareto optimal set in a single run of the algorithm. Additionally, they are not too susceptible to the shape or continuity of the Pareto front (e.g., they can easily deal with discontinuous and concave Pareto fronts).

The first hint regarding the possibility of using evolutionary algorithms to solve a multi-objective problem appears in a Ph.D. thesis from 1967 [35] in which, however, no actual MOEA was developed (the multi-objective problem was restated as a single-objective problem and solved with a genetic algorithm). David Schaffer is normally considered to be the first to have designed an MOEA during the mid-1980s [36]. Schaffer's approach, called Vector Evaluated Genetic Algorithm (VEGA) consists of a simple genetic algorithm with a modified selection mechanism. However, VEGA had a number of problems from which the main one had to do with its inability to retain solutions with acceptable performance, perhaps above average, but not outstanding for any of the objective functions.

After VEGA, the researchers designed a first generation of MOEAs characterized by its simplicity where the main lesson learned was that successful MOEAs had to combine a good mechanism to select non-dominated individuals (perhaps, but not necessarily, based on the concept of Pareto optimality) combined with a good mechanism to maintain diversity (fitness sharing was a choice, but not the only one). The most representative MOEAs of this generation are the following: Nondominated Sorting Genetic Algorithm (NSGA) [37], Niched-Pareto Genetic Algorithm (NPGA) [24] and Multi-Objective Genetic Algorithm (MOGA) [22].

Table 1 Classification of MOEAs

Reference	MOEA	1^{st} Gen.	2^{nd} Gen.
[22]	MOGA	$\surd$	
[24]	NPGA	$\surd$	
[37]	NSGA	$\surd$	
[11]	micro-GA		$\surd$
[19]	NPGA 2		$\surd$
[18]	NSGA-II		$\surd$
[32]	PAES		$\surd$
[15,16]	PESA & PESA-II		$\surd$
[43,44]	SPEA & SPEA2		$\surd$

A second generation of MOEAs started when elitism became a standard mechanism. In fact, the use of elitism is a theoretical requirement in order to guarantee convergence of an MOEA. Many MOEAs have been proposed during the second generation (which we are still living today). However, most researchers will agree that few of these approaches have been adopted as a reference or have been used by others. In this way, SPEA2 and NSGA-II can be considered as the most representative MOEAs of the second generation, also being of interest some others as the Pareto Archived Evolution Strategy (PAES) [32]. Table 1 shows a summary of the most representative MOEAs of both generations.

Finally, we have to point out that nowadays NSGA-II is the paradigm within the MOEA research community since the powerful crowding operator that this algorithm uses usually allows to obtain the widest Pareto sets in a great variety of problems, which is a very appreciated property in this framework. In this way, the question is: "Is NSGA-II the best MOEA to get the desired interpretability-accuracy trade-off of FRBSs following our concrete approach?" (tuning and rule selection). In this work, we analyze the behavior of this algorithm, SPEA2 and two versions of NSGA-II developed to find knees [6] in the optimal Pareto front, which could be a better way to find still accurate solutions but presenting the least possible number of rules. Additionally, we consider two algorithms based on SPEA2 that are specifically designed to address our problem. Next section presents the state-of-the-art on the use of the MOEAs to get this difficult trade-off in order to see how different researchers have faced this problem.

3 Use of MOEAs to Get the Interpretability-Accuracy Trade-off of FRBSs

As mentioned, MOEAs generate a family of equally valid solutions, where each solution tends to satisfy a criterion to a higher extent than another. For this reason, MOEAs have been also applied to improve the difficult trade-off between interpretability and accuracy of FRBSs, where each solution in the Pareto front represents a different trade-off between interpretability and accuracy (see Figure 1).

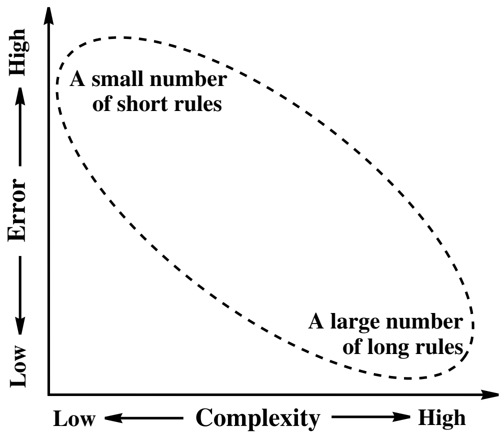


Fig. 1 Trade-off between the error and the interpretability of rule sets

The most continuous and prolific research activity in the application of MOEAs to Mamdani FRBS generation for finding the accuracy-interpretability trade off has been certainly performed by Ishibuchi's group. Earlier works [26] were devoted to the application of simple MOEAs of the first generation to perform a rule selection on an initial set of classification rules involving *"don't care"* conditions and considering two different objectives (classification accuracy and number of rules). Then, a third objective was also included in order to minimize the length of the rules by rule selection [27] or rule learning [27]. In [29], they apply a better MOEA, the Multi-Objective Genetic Local Search [25] (MOGLS), following the same approach for rule selection with three objectives. And finally, two algorithms based on an MOEA of the second generation (NSGA-II) have been proposed respectively for rule selection [33] and rule learning [30] considering the same concepts. In the literature, we can also find some papers of other researchers in this topic. For instance in [14], Cordon et al. use MOGA for jointly performing feature selection and fuzzy set granularity learning with only two objectives.

At this point, we can see that all the methods mentioned were applied to classification problems for rule selection or rule learning, without learning or tuning the MFs that were initially fixed. Most of the works in this topic only consider quantitative measures of the system complexity in order to improve the interpretability of such systems, rarely considering qualitative measures. Moreover, MOEAs considered were slight modifications of MOEAs proposed for general use (MOGA, NSGA-II, etc.) or based on them. Notice that, although NSGA-II improves the results with respect to other MOEAs, since to cross non-dominated rule sets with very different numbers of rules and different rule structures (forced by NSGA-II crowding operator) usually gives a bad accuracy, this MOEA could need of an adaptation to favor the cross of similar solutions in order to also get good results for the accuracy objective (see [33]). A possibility could be the use of similarity measures as the work in [33] (by also favoring the crossover of similar solutions), and other possibility could be to modify the crowding measure as the work in [6] to find knees in multi-objective optimization problems.

On the other hand, there are a few works in the framework of fuzzy modeling for regression problems. In [28], authors show how a simple MOGA can be applied to a three-objective optimization problem to obtain Mamdani FRBSs. In [10], an adaptation of the efficient (2+2)PAES [32] has been applied to the identification of Mamdani FRBSs for regression problems by considering two minimization objectives (the system error and the number of variables involved in the antecedent of the obtained rules). Again, these approaches do not consider learning or tuning of parameters. However, a new method was recently proposed in [5] to perform rule selection and parameter tuning of Mamdani FRBSs by establishing mating restrictions to concentrate the search in the most promising solutions, allowing exploration at first and favoring the exploitation of the most accurate solutions at the end (SPEA2$_{Acc}$). This last approach will be also analyzed and described in this contribution.

Some applications of MOEAs have been also discussed in the literature to improve the difficult trade-off between accuracy and interpretability of Takagi-Sugeno models [38]. In [31, 40, 41], accuracy, interpretability and compactness have been considered as objectives to obtain interpretable and very accurate Takagi-Sugeno models. However, since Takagi-Sugeno models have a linear function in the consequent part of each fuzzy rule, they are close to accuracy representing another type of trade-off with less interpretable models [28]. For this reason, the type of rule most used to achieve the trade-off between accuracy and complexity are the fuzzy rules with linguistic terms in both the antecedent and consequent parts, i.e., Mamdani rules [34].

4 Six Different MOEAs for Rule Selection and Tuning of Membership Functions

As we explain in the previous section most works in the field of fuzzy systems are applied to classification problems by learning or selecting rules, not considering tuning of the MF parameters. The main reason of this fact is that a tuning of parameters implies a lost of the interpretability to some degree. However, it is known that this way to work greatly improves the performance of the linguistic models so obtained, being another alternative to improve the interpretability-accuracy trade-off. For this reason, we would like to show six examples of applications that focus the research in the linguistic fuzzy modeling area, in order to evaluate the performance of MOEAs in a field which is still less explored, and with the objective of inject some ideas or recommendations in this open topic (improvement of the interpretability of very accurate models).

The proposed algorithms will perform rule selection from a given fuzzy rule set together with a parametric tuning of the MFs. To do that, we apply the most used multi-objective algorithms of the second generation, SPEA2 [44] and NSGA-II [18], and two versions of NSGA-II [6] for finding knees. Moreover we consider two extended MOEAs for specific application to this concrete problem, SPEA2$_{Acc}$ in [5], and an extension of that, called SPEA2$_{Acc^2}$. All of them consider two different objectives, system error and number of rules.

In the next subsections, we present SPEA2, NSGA-II, NSGA-II$_A$, NSGA-II$_U$ and SPEA2$_{Acc}$ algorithms and we propose SPEA2$_{Acc^2}$ applied for linguistic fuzzy modeling. At first, the common components of these algorithms are proposed and then the main steps and characteristic of them are described.

4.1 Main Components of the Algorithms

As mentioned, we propose six algorithms to perform rule selection and tuning of MFs and with the aim of improving the desired trade-off between interpretability and accuracy. In the following, the common components needed to apply these algorithms in this concrete problem are explained. They are coding scheme, initial gene pool, objectives and genetic operators:

- **Coding scheme and initial gene pool**
  A double coding scheme for both *rule selection* (C_S) and *tuning* (C_T) is used:

  $$C^p = C_S^p C_T^p$$

  In the $C_S^p = (c_{S1}, \ldots, c_{Sm})$ part, the coding scheme consists of binary-coded strings with size m (with m being the number of initial rules). Depending on whether a rule is selected or not, values '1' or '0' are respectively assigned to the corresponding gene. In the C_T part, a real coding is considered, being m^i the number of labels of each of the n variables comprising the data base,

  $$C_i = (a_1^i, b_1^i, c_1^i, \ldots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), \;\; i = 1, \ldots, n \;\; ,$$
  $$C_T^p = C_1 C_2 \ldots C_n \;\; .$$

  The initial population is obtained with all individuals having all genes with value '1' in the C_S part. And in the C_T part the initial data base is included as first individual. The remaining individuals are generated at random within the corresponding variation intervals. Such intervals are calculated from the initial data base. For each MF, $C_i^j = (a^j, b^j, c^j)$, the variation intervals are calculated in the following way:

  $$[I_{a^j}^l, I_{a^j}^r] = [a^j - (b^j - a^j)/2, a^j + (b^j - a^j)/2]$$
  $$[I_{b^j}^l, I_{b^j}^r] = [b^j - (b^j - a^j)/2, b^j + (c^j - b^j)/2]$$
  $$[I_{c^j}^l, I_{c^j}^r] = [c^j - (c^j - b^j)/2, c^j + (c^j - b^j)/2]$$

- **Objectives**
  Two objectives are minimized for this problem: the number of rules (interpretability) and the Mean Squared Error (accuracy),

  $$\mathrm{MSE} = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2,$$

  with $|E|$ being the size of a data set E, $F(x^l)$ being the output obtained from the FRBS decoded from the said chromosome when the l-th example is considered and y^l being the known desired output. The fuzzy inference system considered to obtain $F(x^l)$ is the *center of gravity weighted by the matching* strategy as defuzzification operator and the *minimum t-norm* as implication and conjunctive operators.

- **Genetic Operators**
  The crossover operator depends on the chromosome part where it is applied: the BLX-0.5 [21] in the C_T part and the HUX [20] in the C_S part.
  Finally, four offspring are generated by combining the two from the C_S part with the two from the C_T part (the two best replace to their parent). The mutation operator changes a gene value at random in the C_S and C_T parts (one in each part) with probability P_m.

- **Importance of the Initial Population**
  Besides, we have to highlight that the way to create the solutions of the initial population for the part of rule selection is a very important factor. Usually, a genetic algorithm generates the initial population totally at random (random selection of the initial rules). However, in this case, to get solutions with a high accuracy we should not lose rules that could present a positive cooperation once their MF parameters have been evolved. The best way to do this is to start with solutions selecting all the possible rules which favors a progressive extraction of bad rules (those that do not improve with the tuning of parameters), only by means of the mutation at the beginning and then by means of the crossover. Different proofs were performed considering a completely random initialization, obtaining simpler solutions but with really worse error values in training and test.

4.2 SPEA2 Based Approach

SPEA2 algorithm [44] was designed to overcome the problems of its predecessor for general multi-objective optimization, SPEA algorithm [43]. In contrast with SPEA,

SPEA2: (1) incorporates a fine-grained fitness assignment strategy which takes into account for each individual the number of individuals that it dominates and the number of individuals by which it is dominated; (2) uses the nearest neighbour density estimation technique which guides the search more efficiently; (3) has an enhanced archive truncation method which guarantees the preservation of boundary solutions. Next, we briefly describe the complete SPEA2 algorithm.

SPEA2 uses a fixed population and archive size. The population forms the current base of possible solutions, while the archive contains the current solutions. The archive is constructed and updated by copying all non-dominated individuals in both archive and population into a temporary archive. If the size of this temporary archive differs from the desired archive size, individuals are either removed or added as necessary. Individuals are added by selecting the best dominated individuals, while the removal process uses a heuristic clustering routine in the objective space. The motivation for this is that one would like to try to ensure that the archive contents represent distinct parts of the objective space. Finally, when selecting individuals for participating in the next generation all candidates are selected from the archive using a binary tournament selection scheme.

Considering the components defined and the descriptions of the authors in [44], SPEA2 algorithm consists of the next steps:

Input: N (population size),
$\overline{N}$ (external population size),
T (maximum number of generations).
Output: A (non-dominated set).

1. Generate an initial population P_0 and create the empty external population $\overline{P}_0 = \emptyset$.
2. Calculate fitness values of individuals in P_t and $\overline{P}_t$.
3. Copy all non-dominated individuals in $P_t \cup \overline{P}_t$ to $\overline{P}_{t+1}$. If $|\overline{P}_{t+1}| > \overline{N}$ apply truncation operator. If $|\overline{P}_{t+1}| < \overline{N}$ fill with dominated in $P_t \cup \overline{P}_t$.
4. If $t \geq T$, return A and stop.
5. Perform binary tournament selection with replacement on $\overline{P}_{t+1}$ in order to fill the mating pool.
6. Apply recombination (BLX-HUX) and mutation operators to the mating pool and set P_{t+1} to the resulting population. Go to step 2 with $t = t + 1$.

4.3 NSGA-II Based Approach

NSGA-II algorithm [18] is one of the most well-known and frequently-used MOEAs for general multi-objective optimization in the literature. As in other evolutionary algorithms, first NSGA-II generates an initial population. Then an offspring population is generated from the current population by selection, crossover and mutation. The next population is constructed from the current and offspring populations. The generation of an offspring population and the construction of the next population are iterated until a stopping condition is satisfied. NSGA-II algorithm has two features, which make it a high-performance MOEA. One is the fitness evaluation of each solution based on Pareto ranking and a crowding measure, and the other is an elitist generation update procedure.

Each solution in the current population is evaluated in the following manner. First, Rank 1 is assigned to all non-dominated solutions in the current population. All solutions with Rank 1 are tentatively removed from the current population. Next, Rank 2 is assigned to all non-dominated solutions in the reduced current population. All solutions with Rank 2 are tentatively removed from the reduced current population. This procedure is iterated until all solutions are tentatively removed from the current population (i.e., until ranks are assigned to all solutions). As a result, a different rank is assigned to each solution. Solutions with smaller ranks are viewed as being better than those with larger ranks. Among solutions with the same rank, an additional criterion called a crowding measure is taken into account.

The crowding measure for a solution calculates the distance between its adjacent solutions with the same rank in the objective space. Less crowded solutions with larger values of the crowding measure are viewed as being better than more crowded solutions with smaller values of the crowding measure.

A pair of parent solutions are selected from the current population by binary tournament selection based on the Pareto ranking and the crowding measure. When the next population is to be constructed, the current and offspring populations are combined into a merged population. Each solution in the merged population is evaluated in the same manner as in the selection phase of parent solutions using the Pareto ranking and the crowding measure. The next population is constructed by choosing a specified number (i.e., population size) of the best solutions from the merged population. Elitism is implemented in NSGA-II algorithm in this manner.

Considering the components previously defined and the descriptions of the authors in [18], NSGA-II consists of the next steps:

1. A combined population R_t is formed with the initial parent population P_t and offspring population Q_t (initially empty).
2. Generate all non-dominated fronts $F = (F_1, F_2, ...)$ of R_t.
3. Initialize $P_{t+1} = 0$ and $i = 1$.
4. Repeat until the parent population is filled.
5. Calculate crowding-distance in F_i.
6. Include i-th non-dominated front in the parent population.
7. Check the next front for inclusion.
8. Sort in descending order using crowded-comparison operator.

9. Choose the first $(N - |P_{t+1}|)$ elements of F_i.
10. Use selection, crossover (BLX-HUX) and mutation to create a new population Q_{t+1}.
11. Increment the generation counter.

4.4 NSGA-II with Angle-Measure Based Approach

As mentioned, the performance of NSGA-II relies on two measures when comparing individuals: The first is the non-domination rank and, if two individuals have the same non-domination rank, as a secondary criterion, a crowding measure is used.

In [6], authors presented a different version of NSGA-II in order to find knees in the Pareto front by modifying the secondary criterion, and replacing the crowding measure by either an angle-based measure or an utility-based measure. Again, this algorithm was proposed for multi-objective optimization in general. However, in our case, a knee could represent the best compromise between accuracy and number of rules. So we propose the use of these kinds of measures to search for these interesting Pareto zones in our concrete problem. In this subsection, the use of the angle-based measure is explained in order to replace the crowding measure of NSGA-II.

In the case of only two objectives, the trade-offs in either direction can be estimated by the slopes of the two lines through an individual and its two neighbors. The angle between these slopes can be regarded as an indication of whether the individual is at a knee or not. For an illustration, consider Figure 2. Clearly, the larger the angle α between the lines, the worse the trade-offs in either direction, and the more clearly the solution can be classified as a knee.

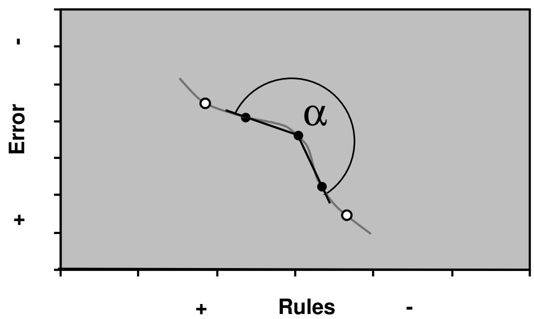


Fig. 2 Calculation of the angle measure

More formally, to calculate the angle measure for a particular individual C^i, we calculate the angle between the individual and its two neighbors, i.e. between (C^{i-1}, C^i) and (C^i, C^{i+1}). These three individuals have to be pairwise linearly independent, thus duplicate individuals (individuals with the same objective function values, which are not prevented in NSGA-II per se) are treated as one and are assigned the same angle-measure. If no neighbor to the left (right) is found, a horizontal (vertical) line is used to calculate the angle. Similar to the standard crowding measure, individuals with a larger angle-measure are preferred.

Calculating the angle measure in 2D is efficient. For more than two objectives, however, it becomes impractical even to just find the neighbors. Thus, we restrict our examination of the angle-based focus to problems with two objectives only. Another important issue is that the values of the different objectives have to be normalized in order to calculate fair angle values. In our case, the sides of the triangles used to compute the final value of α are divided by the difference between the best and the worst values of the corresponding objective in the current Pareto front, as it is done in the original NSGA-II to compute the crowding measure.

4.5 NSGA-II with Utility-Measure Based Approach

An alternative measure for a solution's relevance was also proposed in [6]. This subsection explains the use of this measure (utility-based measure) in order to provide a different way to replace the crowding measure of NSGA-II.

The proposed alternative measure is the expected marginal utility that a solution provides to a decision maker, assuming linear utility functions of the form $U(C, \lambda) = \lambda f_1(C) + (1 - \lambda) f_2(C)$, with all $\lambda \in [0, 1]$ being equally likely. For illustration, let us first assume we would know that the decision maker has a particular preference function $U(C, \lambda')$, with some known λ'. Then, we could calculate, for each individual C^i in the population, the decision maker's utility $U(C^i, \lambda')$ of that individual. Clearly, given the choice among all individuals in the population, the decision maker would select the one with the highest utility. Now let us define an individual's marginal utility $U'(C, \lambda')$ as the additional cost the decision maker would have to accept if that particular individual would not be available and he/she would have to settle for the second best, i.e.

$$U'(C^i, \lambda') = \begin{cases} \min_{j \neq i} U(C^j, \lambda') - U(C^i, \lambda') & \\ & : i = \arg\min U(C^j, \lambda') \\ 0 & : otherwise \end{cases}$$

The proposed utility measure assumes a distribution of utility functions uniform in the parameter λ in order to calculate the expected marginal utility. For the case of only two objectives, the expected marginal utility can be calculated exactly by integrating over all possible linear utility functions. However, the expected marginal utilities can be approximated simply by sampling, i.e. by calculating the marginal utility for all individuals for a number of randomly chosen utility functions, and taking the average as expected marginal utility. Sampling can be done either randomly or, as was proposed in [6] in order to reduce variance, in a systematic manner (equidistant values for λ). The number of utility functions used for approximation was called precision of the measure. Authors recommend a precision of at least the number of individuals in the population. Naturally, individuals with the largest overall marginal utility are preferred.

Notice, however, that the assumption of linear utility functions makes it impossible to find knees in concave regions of the non-dominated front. Unlike the angle measure, the utility measure extends easily to more than two objectives, by defining $U(C, \lambda) = \Sigma \lambda_i f_i(C) \ with \ \Sigma \lambda_i = 1$.

In this paper, the marginal utilities have been computed by sampling, considering equi-distant values for λ and a precision of exactly the number of individuals in the population. As in the case of the angle-measure based approach, the values of the different objectives have to be normalized in order to calculate fair utility values. In our case, objective values considered for computing utility values were normalized considering the best and the worst values of the corresponding objective in the current Pareto front. In this way, an objective value can be normalized as,

$$f_i^{'}(C) = (f_i(C) - f_i^{MIN})/f_i^{MAX} \quad ,$$

providing values between 0.0 and 1.0.

4.6 Accuracy-Oriented Based Approach: SPEA2$_{Acc}$ Algorithm

SPEA2$_{Acc}$ algorithm was very recently proposed in [5], and is a particularization of SPEA2 based approach presented in Section 4.2 to better solve the problem of rule selection and tuning of FRBSs. This algorithm tries to focus the search on the desired Pareto zone, high accuracy with least possible number of rules, proposing two main changes on SPEA2 algorithm with the aim of giving more selective pressure to those solutions that have a high accuracy (crossing dissimilar solutions in principle and similar ones at the end). These changes were also applied and analyzed on NSGA-II in [5] showing not so good results. The proposed changes are described next:

- A restarting operator is applied exactly at the mid of the algorithm, by maintaining the most accurate individual as the sole individual in the external population ($\overline{P}_{t+1}$ with size 1) and obtaining the remaining individuals in the population (P_{t+1}) with the same rule configuration of the best individual and tuning parameters generated at random within their corresponding variation intervals. This operation is performed in step 4 (see Section 4.2) as a second condition, then returning to step 2 with $t = t + 1$. In this way, the search is concentrated only in the desired Pareto zone (similar solutions in a zone with high accuracy).
- In each stage of the algorithm (before and after restarting), the number of solutions in the external population ($\overline{P}_{t+1}$) considered to form the mating pool is progressively reduced, by focusing only on those with the best accuracy. To do that, the solutions are sorted from the best to the worst (considering accuracy as sorting criterion) and the number of solutions considered for selection is reduced progressively from 100% at the beginning to 50% at the end of each stage.

4.7 Extension of SPEA2$_{Acc}$ Algorithm: SPEA2$_{Acc^2}$

SPEA2$_{Acc}$ algorithm tries to focus the search in the Pareto zone containing the most accurate solutions. This algorithm represents a good way to obtain more accurate solutions by maintaining only a few more rules with respect to its counterpart (SPEA2). However, sometimes this fact could represent a problem since there are problems in which to obtain accurate solutions could be easy but not so easy to remove unnecessary rules. In this subsection, we propose an extension of this algorithm in order to solve this problem. To do that, we propose two changes based on our experience in this concrete problem:

- An intelligent crossover that is able to profit even more from the corresponding parents, replacing the HUX crossover for the C_S part. To obtain each offspring the following steps are applied:
  1. The BLX crossover is applied to obtain the C_T part of the offspring.
  2. Once the real parameters are obtained determining a the data base, for each gene in the C_S part the corresponding rule is independently extracted from each individual involved in the crossover (offspring and parents 1 and 2). In this way, the same rule is obtained three times with different MFs (those concerning these three individuals).
  3. Euclidean normalized distances are computed between offspring and each parent by only considering the center points (vertex) of the MFs involved in the extracted rules. The differences between each two points are normalized by the amplitude of their respective variation intervals.
  4. The nearest parent is the one that determines if this rule is selected or not for the offspring by directly copying its value in C_S for the corresponding gene.
  5. This process is repeated until all the C_S values are assigned for the offspring.

  Four offspring are obtained repeating this process four times (after considering mutation, only the two most accurate are taken as descendent). By applying this operator, exploration is performed in the C_T part and the C_S part is directly obtained based on the previous knowledge each parent has about the use or not of a specific configuration of MFs for each rule. This avoid to recover a bad rule that was discarded for a concrete configuration of MFs, or allow to recover a good rule that is still considered for a

concrete configuration of MFs, increasing the probability of succeed in the selection or elimination of a rule for each concrete configuration of MFs.
- Since a better exploration is performed for the C_S part, the mutation operator does not need to add rules (rules that were eliminated in the parents for a similar bad configuration of the MFs involved in these rules). In this way, once an offspring is generated the mutation operator changes a gene value at random in the C_T part (as in the previous algorithm) and directly sets to zero a gene selected at random in the C_S part (one gene is considered in each part) with probability P_m.

Applying these operators two problems are solved. Firstly, crossing individuals with very different rule configurations is more productive. And secondly, this way to work favors rule extraction since mutation is only engaged to remove unnecessary rules.

5 Experiments

To evaluate the goodness of the proposed approaches, two real-world problems with different complexities (different number of variables and available data) are considered to be solved:

- An *electrical distribution problem* [13] that consists of estimating the maintenance costs of medium voltage lines in a town (1059 cases; 4 continuous variables).
- The *Abalone* dataset [42] that concerns the task of trying to predict the number of rings in the shells of abalone (which is related to their age) based on a series of biometric measures of these animals (4177 cases; 7 continuous variables; 1 nominal variable).

Methods considered for the experiments are briefly described in Table 2. In both problems, WM method is considered to obtain the initial set of fuzzy rules. To do so, we will consider symmetrical fuzzy partitions of triangular-shaped MFs. Once the initial rule set is generated, the proposed post-processing algorithms will be applied. T and S methods perform the tuning of parameters and rule selection respectively. TS indicates tuning together with rule selection in the same algorithm. All of them consider the accuracy of the model as the sole objective. MOEAs studied in this work (TS-SPEA2, TS-NSGA-II, TS-NSGA-II$_A$, TS-NSGA-II$_U$, TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$) perform rule selection from a given fuzzy rule set together with the parametric tuning of the MFs considering two objectives, system error and number of rules.

In the next subsections, the named problems are introduced and solved to analyze the behavior of the proposed methods. To do that, the experimental set-up is first described. Finally, at the end of this section an internal study on alternative possibilities to select solutions from final Paretos and on the initialization influence is

Table 2 Methods Considered for Comparison

Méthod	Ref.	Description
WM	[39]	Wang & Mendel algorithm
T	[9]	Tuning of Parameters
S	[9]	Rule Selection
TS	[9]	Tuning & Selection
Application of standard MOEAs for general use		
TS-SPEA2	[5]	Tuning & Selection by SPEA2
TS-NSGA-II	[5]	Tuning & Selection by NSGA-II
TS-NSGA-II$_A$		Tuning & Selection by NSGA-II$_{angle}$
TS-NSGA-II$_U$		Tuning & Selection by NSGA-II$_{utility}$
Extended MOEAs for specific application		
TS-SPEA2$_{Acc}$	[5]	Accuracy-Oriented SPEA2
TS-SPEA2$_{Acc^2}$		Extension of SPEA2$_{Acc}$

also performed (considering the initialization presented in Section 4.1 or a completely random initialization).

5.1 Experimental Set-Up

To develop the different experiments, we consider a *5-folder cross-validation model*, i.e., 5 random partitions of data each with 20%, and the combination of 4 of them (80%) as training and the remaining one as test. For each one of the 5 data partitions, the post-processing methods have been run 6 times (6 different seeds), showing for each problem the averaged results of a total of 30 runs.

In the case of methods with multi-objective nature (TS-SPEA2, TS-NSGA-II, TS-NSGA-II$_A$, TS-NSGA-II$_U$, TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$), the averaged values are calculated considering the most accurate solution from each Pareto obtained. In this way, the multi-objective algorithms can be compared with several single objective based methods. This way to work differs with the previous works in the specialized literature (see section 3) in which one or several Pareto fronts are presented and an expert should after select one solution. Our main aim following this approach is to compare the same technique when only the accuracy objective is considered (algorithm WM+TS) with the most accurate solution found by the proposed multi-objective algorithms in order to see if the Pareto fronts obtained are not only wide but also optimal (almost similar solutions to that obtained by WM+TS should be included in the final Pareto).

The values of the input parameters considered by S, T and TS (single objective oriented algorithms) are[1]: population size of 61, 100000 evaluations, 0.6 as crossover probability and 0.2 as mutation probability per chromosome. In the case of MOEAs, the most important param-

[1] With these values we have tried to ease the comparisons selecting standard common parameters that work well in most cases instead of searching for very specific values to each method. Moreover, we have set a large number of evaluations in order to allow the compared algorithms to achieve an appropriate convergence.

eter is the population size. In the case of SPEA2 based algorithms, a good proportion between standard population and external population is 3/1 or 4/1. Different population sizes were probed showing not very different results but presenting the best performance when the external population took values between 50 and 100 individuals. It is, when the population used for parent selection has similar sizes than those considered by single objective oriented algorithms in these kinds of problems. In this way, we have considered an external population size of 61 (the same size used by the named algorithms with single objective) and a proportion of 1/3 rounded to 200 as standard population size. In the case of NSGA-II based algorithms, since the archive and the population have the same size and they have been usually used with values of 200 and 100 as population size in general problems for continuous optimization, a good value could be 200 as population size (the same that SPEA2 based approaches). However, although 100, 200 and 400 presented a similar and reasonable performance for these algorithms, the best results were obtained by taking similar sizes than those considered by S, T and TS (with single objective) in these kinds of problems (i.e., when the size of the population used for parent selection takes these values). Therefore, we recommend the use of this simple rule of thumb to fix the population size in these kinds of problems. Finally, the values of the input parameters considered by the MOEAs are shown in the next: population size of 200 (61 in the case of NSGA-II based algorithms), external population size of 61 (in the case of SPEA2 based algorithms), 100000 evaluations and $P_m = 0.2$ as mutation probability per chromosome.

5.2 Estimating the Maintenance Costs of Medium Voltage Lines

Estimating the maintenance costs of the medium voltage electrical network in a town [13] is a complex but interesting problem. Since a direct measure is very difficult to obtain, it is useful to consider models. These estimations allow electrical companies to justify their expenses. Moreover, the model must be able to explain how a specific value is computed for a certain town. Our objective will be to relate the *maintenance costs of the medium voltage lines* with the following four variables: *sum of the lengths of all streets in the town*, *total area of the town*, *area that is occupied by buildings*, and *energy supply to the town*. We will deal with estimations of minimum maintenance costs based on a model of the optimal electrical network for a town in a sample of 1,059 towns. As said, five data partitions[2] considering an 80% (847) in training and a 20% (212) in test are considered for the experiments. The initial linguistic partitions are comprised by *five linguistic terms* with equally distributed triangular shaped MFs.

Table 3 Results obtained by the studied methods in the electrical distribution problem

Method	**#R**	**MSE**$_{tra}$	σ_{tra}	**t**	**MSE**$_{tst}$	σ_{tst}	**t**
WM	65.0	57605	2841	+	57934	4733	+
WM+T	65.0	17020	1893	+	21027	4225	+
WM+S	40.9	41158	1167	+	42988	4441	+
WM+TS	41.3	13387	1153	+	17784	3344	+
TS-SPEA2	**28.9**	11630	1283	+	15387	3108	=‡
TS-NSGA-II	31.4	11826	1354	+	16047	4070	+
TS-NSGA-II$_A$	29.7	11798	1615	+	16156	4091	+
TS-NSGA-II$_U$	30.7	11954	1768	+	15879	4866	+
TS-SPEA2$_{Acc}$	32.3	10714	1392	=	14252	3181	=
TS-SPEA2$_{Acc^2}$	29.8	**10325**	**1121**	*	**13935**	**2759**	*

‡ + with 94% confidence

The results obtained by the analyzed methods are shown in Table 3, where $\#R$ stands for the number of rules, MSE$_{tra}$ and MSE$_{tst}$ respectively for the averaged error obtained over the training and test data, σ for the standard deviation and t for the results of applying a *student's t-test* (with 95 percent confidence) in order to ascertain whether differences in the performance of the best results are significant when compared with that of the other algorithms in the table. The interpretation of this column is:

- $\star$ represents the best averaged result.
- + means that the best result has better performance than that of the corresponding row.

Analysing the results showed in Table 3 we can highlight the following facts:

- Methods based on SPEA2 show a reduction of MSE$_{tra}$ and MSE$_{tst}$ with respect to the models obtained by only considering the accuracy objective (WM+TS). Moreover, a considerable number of rules have been removed from the initial FRBS, obtaining simpler models with a better performance.
- NSGA-II based algorithms statistically obtain the same accuracy than the models obtained with TS-SPEA2 considering the most accurate result of each obtained Pareto. However, all of them present a higher number of rules (from one to three) and worse average values than TS-SPEA2. Moreover, a difference with TS-SPEA2 is that comparing each of the NSGA-II based approaches with WM+TS (single objective-based approach) the *student's t-test* would show that they are statistically equal in their generalization ability (MSE$_{tst}$). Therefore, we could consider that these algorithms get good solutions, since in any case, they are quite similar to the application of the standard SPEA2 (specially in the case of TS-NSGA-II$_A$ and TS-NSGA-II$_U$).
- The best results were obtained by TS-SPEA2$_{Acc^2}$ and TS-SPEA2$_{Acc}$, showing that the use of expert

[2] These data sets are available at: *http://decsai.ugr.es/∼casillas/fmlib/*.

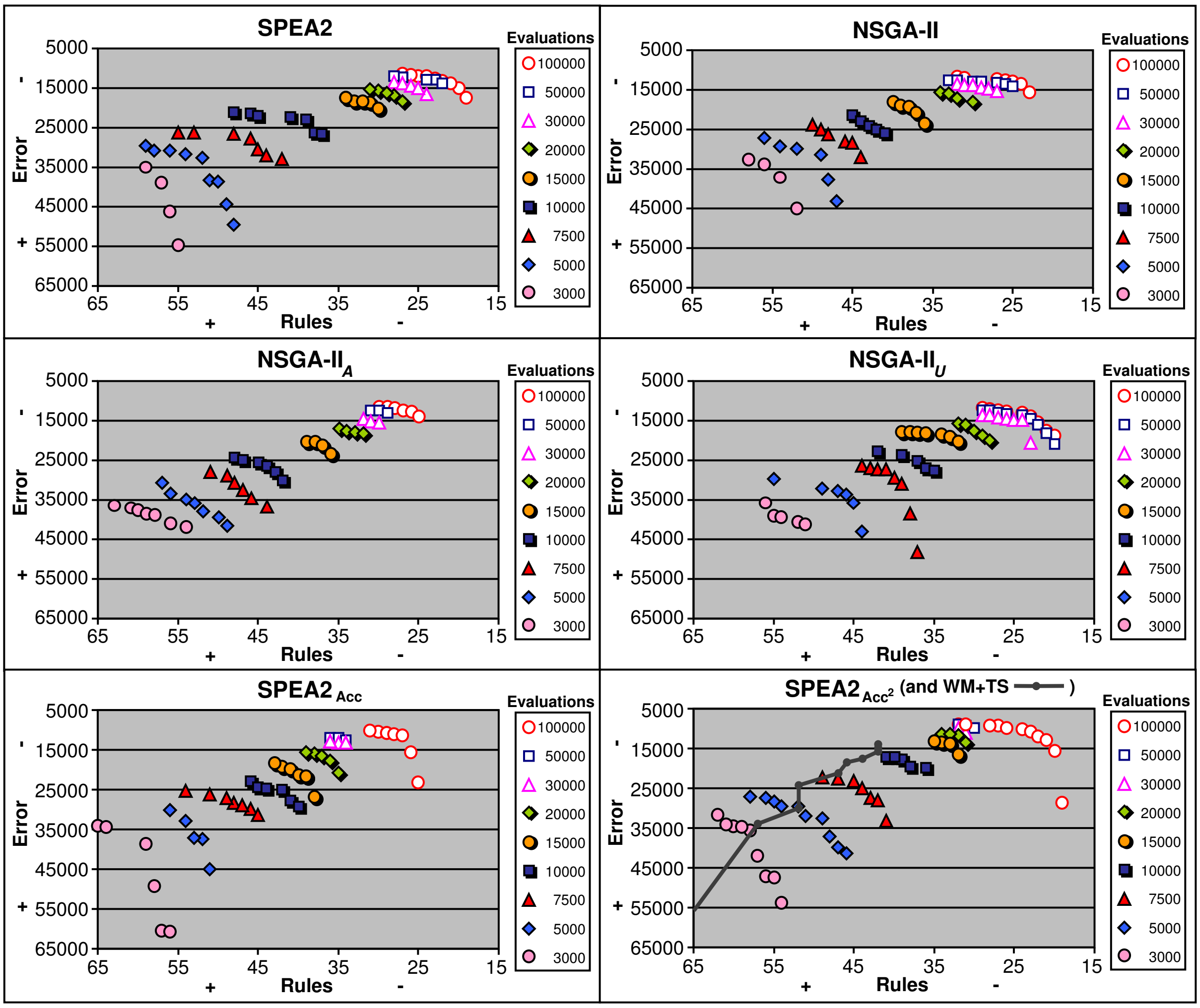


Fig. 3 Example of the Pareto front evolution along one representative run of TS-SPEA2, TS-NSGA-II, TS-NSGA-II$_A$, TS-NSGA-II$_U$, TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$ in the Electrical distribution problem. Evolution of the best solution in a representative run with WM+TS is also included together with TS-SPEA2$_{Acc^2}$

knowledge in the design process can help to obtain more optimal Pareto fronts. Moreover, TS-SPEA2$_{Acc^2}$ is able to obtain the best average values with even less rules than the TS-SPEA2$_{Acc}$ algorithm.

All MOEAs considered obtain significantly simpler models that those obtained by only considering the accuracy based objective and almost the same results (presenting minor average values in all the cases and statistical differences in the case of the extended MOEAs). This is a positive fact since an appropriate use of MOEAs can improve the desired trade-off with respect to the classic accuracy-based approaches, and specific adaptations can help to improve the performance of standard MOEAs.

These results (more simple and accurate models by applying a multi-objective approach) are due to the large search space that involves these kinds of problems. There are some initial rules that should be removed since they do not cooperate in a good way with the remaining ones. Even in the case of only considering an accuracy-based objective, the large search space that supposes the tuning of parameters makes very difficult to remove these kinds of rules since bad rules are tuned together with the remaining ones searching for their best cooperation. The use of a multi-objective approach favors a better selection of the ideal number of rules, preserving some rule configurations until the rule parameters are evolved to dominate solutions including bad rules, which can finally lead to solutions with more freedom-degrees to tune the corresponding parameters involving a better cooperation among the different rules.

In Figure 3, we can see the Pareto evolution in a representative run for each multi-objective algorithm and also the evolution of the best solution in the population in a representative run of WM+TS. Each type of symbol in the figure represents the Pareto solutions at different stages of the evolution (caption ‘Evaluations’ shows the number of evaluations in which each Pareto was taken in a simple run and the symbol associated). We can observe as the Pareto moves along without having a wide extension but dominating the solution obtained by WM+TS

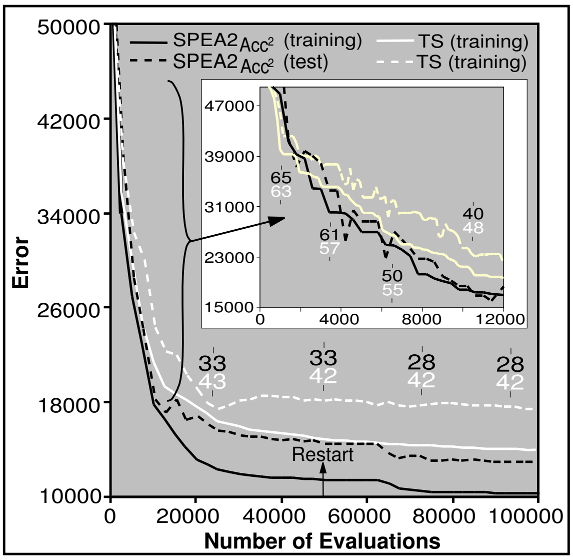


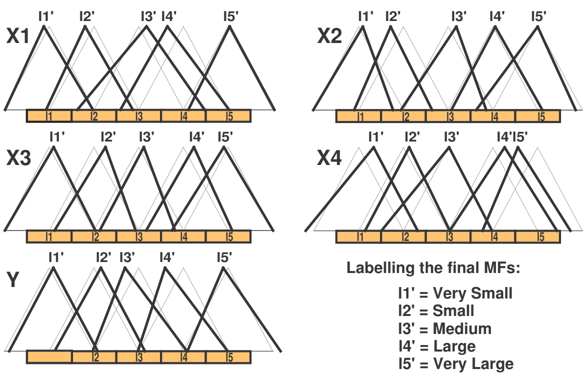
Fig. 4 Convergence in a representative run considering the best solution of the population of WM+TS and the most accurate solution in the Pareto of TS-SPEA2$_{Acc^2}$. Numbers in black and white represent the number of rules in the solution in different moments of the evolution

at the end. Although these figures only represent a run for each algorithm, the obtained Pareto fronts in the different runs (5 fold, six seeds, 30 runs, after 100000 evaluations) are in general very similar to those showed in Figure 3 for each algorithm. In this way, the most accurate solution of each Pareto can be considered as the position in which we can find a sort set of close solutions representing different trade-offs in the Pareto zone with still accurate solutions.

Another important fact is that in the final Pareto fronts of TS-SPEA2$_{Acc^2}$ and TS-SPEA2$_{Acc}$ there are one or two solutions (those with the minor number of rules) showing a bad performance with respect to the remaining ones in their respective Paretos. This situation is typical in practically all the obtained Pareto fronts and in all the approaches considered. These solutions represent new individuals that appeared at the end of the evolution practically without time to evolve their associated MFs in a zone in which to extract a rule without severely affecting the accuracy is more difficult.

Figure 4 shows the convergence of the best solution of the population from WM+TS and the most accurate solution in the Pareto from TS-SPEA2$_{Acc^2}$ in a representative run for the electrical distribution problem. An interesting fact is that the WM+TS (single objective) algorithm is faster at the beginning (in training and number of rules) while TS-SPEA2$_{Acc^2}$ takes a bit more time for exploration in order to take further advantage. Another interesting fact is to see how to perform restarting at the mid of the TS-SPEA2$_{Acc^2}$ process positively affects the values in training and number of rules.

Figures 5 and 6 respectively show the most accurate models obtained with WM+TS and TS-SPEA2$_{Acc^2}$ in the electrical distribution problem. To ease graphic representation, in these figures, the MFs are labeled from $l1$ to lm^i. Nevertheless, such MFs are initially associated to a linguistic meaning determined by an expert. In this way, if the $l1$ label of the $X1$ variable represents 'Very



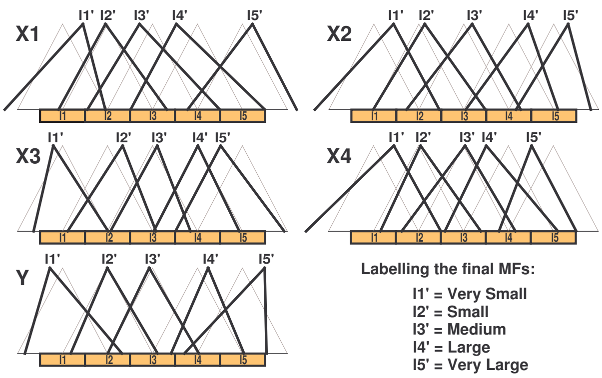
#R: 43 *MSE-tra: 11383* *MSE-tst: 13416*

X1	X2	X3	X4	Y
l1'	l1'	l1'	l1'	l1'
l1'	l1'	l1'	l2'	l2'
l1'	l2'	l1'	l1'	l1'
l1'	l2'	l2'	l2'	l2'
l2'	l1'	l1'	l1'	l1'
l2'	l1'	l1'	l2'	l2'
l2'	l1'	l2'	l1'	l2'
l2'	l1'	l2'	l2'	l2'
l2'	l2'	l1'	l1'	l1'
l2'	l2'	l1'	l2'	l2'
l2'	l2'	l2'	l1'	l2'
l2'	l2'	l2'	l2'	l2'
l2'	l3'	l3'	l2'	l3'
l3'	l2'	l1'	l1'	l1'

X1	X2	X3	X4	Y
l3'	l2'	l1'	l2'	l2'
l3'	l2'	l1'	l3'	l2'
l3'	l2'	l2'	l3'	l3'
l3'	l3'	l2'	l1'	l2'
l3'	l3'	l2'	l2'	l2'
l3'	l3'	l2'	l3'	l3'
l3'	l3'	l3'	l2'	l3'
l3'	l4'	l3'	l2'	l3'
l3'	l4'	l4'	l3'	l4'
l4'	l2'	l2'	l1'	l2'
l4'	l2'	l2'	l2'	l2'
l4'	l3'	l2'	l1'	l2'
l4'	l3'	l2'	l3'	l3'
l4'	l3'	l2'	l4'	l3'

X1	X2	X3	X4	Y
l4'	l3'	l3'	l2'	l3'
l4'	l3'	l3'	l3'	l4'
l4'	l4'	l3'	l1'	l3'
l4'	l4'	l3'	l4'	l4'
l4'	l4'	l4'	l2'	l4'
l4'	l4'	l4'	l4'	l5'
l4'	l5'	l4'	l2'	l3'
l4'	l5'	l4'	l3'	l4'
l4'	l5'	l5'	l2'	l5'
l4'	l5'	l5'	l3'	l5'
l5'	l2'	l2'	l5'	l4'
l5'	l2'	l3'	l2'	l3'
l5'	l2'	l3'	l5'	l4'
l5'	l4'	l3'	l4'	l4'
l5'	l4'	l3'	l5'	l5'

Fig. 5 DB with/without tuning (black/gray) and RB of the best model (in training) obtained by WM+TS in the Electrical distribution problem



#R: 28 *MSE-tra: 8232* *MSE-tst: 14670*

X1	X2	X3	X4	Y
l1'	l1'	l1'	l1'	l1'
l1'	l1'	l1'	l2'	l2'
l2'	l1'	l1'	l1'	l1'
l2'	l1'	l1'	l2'	l2'
l2'	l1'	l2'	l2'	l2'
l2'	l2'	l2'	l1'	l2'
l2'	l3'	l3'	l1'	l3'
l3'	l2'	l1'	l1'	l1'
l3'	l2'	l1'	l2'	l2'

X1	X2	X3	X4	Y
l3'	l2'	l1'	l3'	l2'
l3'	l2'	l2'	l3'	l3'
l3'	l3'	l2'	l2'	l2'
l3'	l3'	l3'	l2'	l3'
l3'	l4'	l3'	l3'	l3'
l4'	l2'	l2'	l2'	l2'
l4'	l3'	l2'	l1'	l2'
l4'	l3'	l2'	l3'	l3'
l4'	l3'	l2'	l4'	l3'

X1	X2	X3	X4	Y
l4'	l3'	l3'	l2'	l3'
l4'	l3'	l3'	l3'	l4'
l4'	l4'	l3'	l2'	l3'
l4'	l4'	l4'	l2'	l4'
l4'	l4'	l4'	l4'	l5'
l4'	l5'	l4'	l2'	l3'
l4'	l5'	l5'	l3'	l5'
l5'	l2'	l2'	l5'	l4'
l5'	l2'	l3'	l2'	l3'
l5'	l4'	l3'	l5'	l5'

Fig. 6 DB with/without tuning (black/gray) and RB of the best model (in training) obtained by TS-SPEA2$_{Acc^2}$ in the Electrical distribution problem

Small', $l1'$ could be also interpreted as 'Very Small' as classically has been considered when tuning is applied or based on the expert opinion, maintaining the original meaning of such label or renaming it if possible. This is the case of Figures 5 and 6, where in principle practically all the new labels could maintain their initial meanings.

5.3 Predicting the Abalone Age

The Abalone dataset [42] is concerned with predicting the age of an Abalone specimen (a type of shellfish) based on physical measurements. Why is it interesting to predict age? For ecologic and commercial fish farming purposes, the age composition of abalone populations are relevant. Here the number of rings is proxy for age. The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope.

However, this is a boring, time-consuming and expensive task. Other measurements, which are easier to obtain, are therefore used to predict the age. Recorded measurements on 4177 Abalone[3], of interest in determining relationships useful to predicting the age of future abalone from easily made physical measurements, were obtained from the Marine Resources Division at the Department of Primary Industry and Fisheries, Tasmania, and can be used for this task. The goal is to predict the *number of rings* based on the following eight variables (seven continuous and one nominal): *sex* (nominal), *length*, *diameter*, *height*, *whole weight*, *shucked weight*, *viscera weight*, and *shell weight*. As explained, five data partitions considering an 80% (3342) in training and a 20% (835) in test are considered for the experiments. In this case, the initial linguistic partitions are comprised by *three linguistic terms* with equally distributed triangular shaped MFs. The accuracy of the models obtained is quite similar by considering three or five linguistic terms in this problem and therefore a number of three labels per variable is preferable since the final models are comprised of a smaller number of rules.

The results obtained in this problem by the analyzed methods are shown in Table 4 (these kinds of table was described in the previous subsection). On the Abalone data set (with very high level of noise), all of the MOEAs and even WM+TS achieved almost identical performance, however they present different numbers of rules. This problem is quite different to that in the previous section. As can be seen, WM+S obtains the models with the smaller number of rules, which indicates that there are a lot of rules that can be removed from the initial system. So, in this problem, the real challenge would be to remove these rules in an appropriate manner instead of trying very important accuracy improvements.

[3] Available from the UCI Machine Learning Repository (www.ics.uci.edu/∼mlearn/MLRepository.html).

Table 4 Results obtained by the studied methods in the Abalone dataset

Method	**#R**	**MSE**$_{tra}$	σ_{tra}	**t**	**MSE**$_{tst}$	σ_{tst}	**t**
WM	68.2	8.407	0.443	+	8.422	0.545	+
WM+T	68.2	2.688	**0.063**	+	2.770	0.242	+
WM+S	**18.0**	4.825	1.078	+	4.795	1.165	+
WM+TS	28.4	2.473	0.097	+	2.582	0.290	=
TS-SPEA2	20.0	2.383	0.078	=	2.518	0.246	=
TS-NSGA-II	22.4	2.398	0.084	=	2.526	0.242	=
TS-NSGA-II$_A$	22.1	2.404	0.098	=	2.535	0.265	=
TS-NSGA-II$_U$	21.8	2.407	0.082	=	2.520	0.237	=
TS-SPEA2$_{Acc}$	22.2	**2.368**	0.085	*	**2.511**	0.263	*
TS-SPEA2$_{Acc^2}$	18.6	2.372	0.075	=	2.517	**0.230**	=

Analyzing the results presented in Table 4 we can stress the following facts:

- In this case, although TS-SPEA2$_{Acc}$ method presents the best average results in MSE$_{tra}$ and MSE$_{tst}$ with respect to the remaining models, TS-SPEA2$_{Acc^2}$ could be considered as the best approach since practically the same values were obtained in training and test, and the best value in number of rules has been obtained.
- In this case, NSGA-II based algorithms are statistically equal than those models obtained by application of the standard SPEA2, but again, all of them present a higher number of rules (about 2). However, accuracy differences are practically not appreciated showing results quite similar to TS-SPEA2. This time, TS-NSGA-II$_A$ and TS-NSGA-II$_U$ obtain more or less the same number of rules than TS-NSGA-II, although their average numbers of rules are still better.
- All MOEAs considered obtain significantly simpler models that those obtained by only considering the accuracy based objective and almost the same results, improving again the desired trade-off with respect to the classic accuracy-based approaches.

In Figure 7, we can see the Pareto evolution in a representative run for each multi-objective algorithm (these kinds of figures were described in the previous subsection). Once more, the different Pareto fronts move along without having a wide extension. TS-NSGA-II$_U$, TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$ show the wider Pareto fronts in their corresponding figures. However, the front obtained by TS-SPEA2$_{Acc^2}$ is again located more to the top right zone (the zone with less rules and more accurate models).

Figure 8 shows the most accurate model obtained with TS-SPEA2$_{Acc^2}$ in the electrical distribution problem (these kinds of figures were also described in the previous subsection). Although practically all the MFs could maintain their original meanings (from a subjective point of view), there are two cases that probably should be renamed by experts if possible, l2' and l3' in X1 and X6 respectively. From a subjective point of

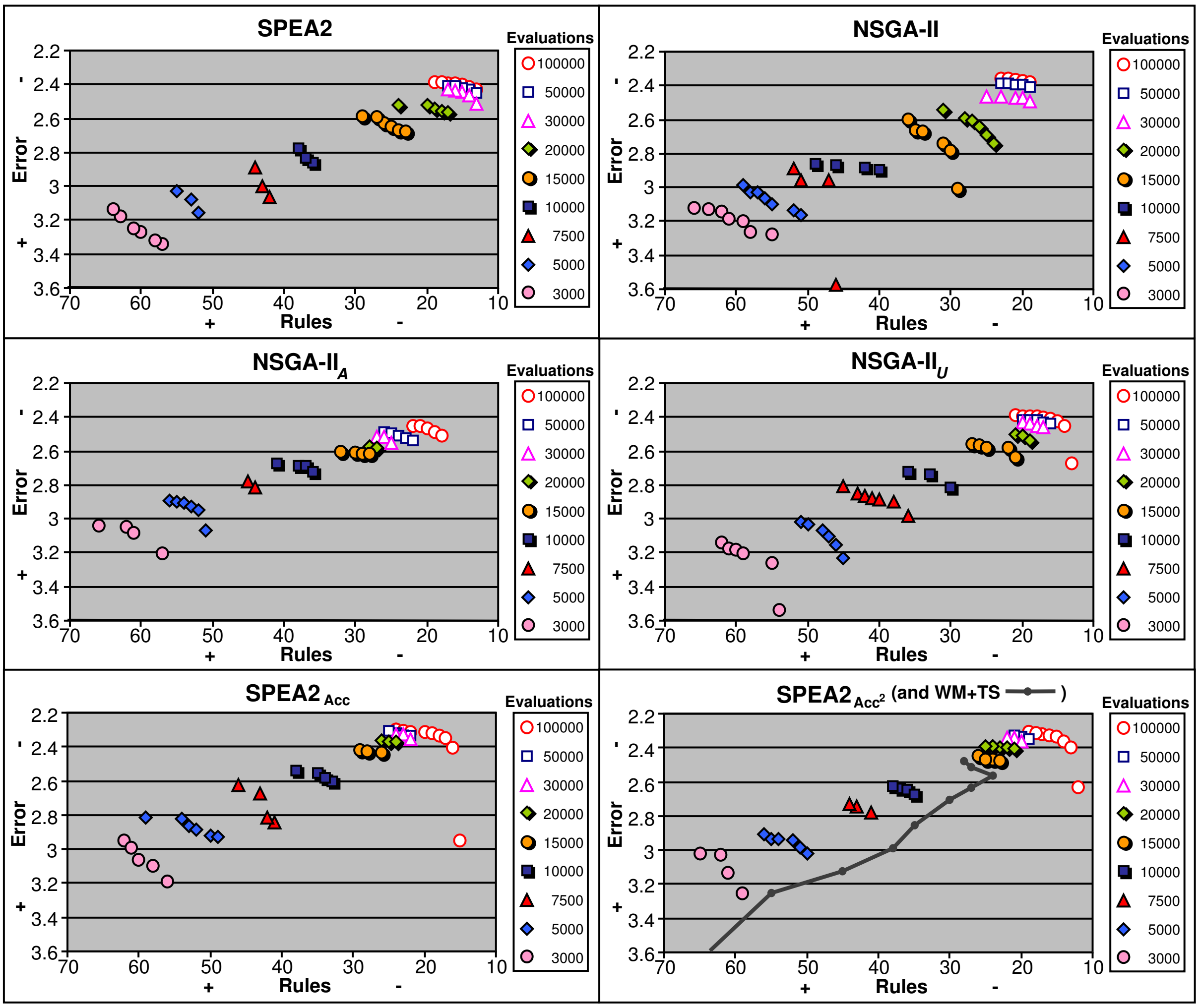


Fig. 7 Example of the Pareto front evolution along one representative run of TS-SPEA2, TS-NSGA-II, TS-NSGA-II$_A$, TS-NSGA-II$_U$, TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$ in the Abalone dataset. Evolution of the best solution in a representative run with WM+TS is also included together with TS-SPEA2$_{Acc^2}$

view we show a way to rename them. The most accurate model obtained from WM+TS has been not included for this problem in order to avoid an excessive length of the paper since, as in the previous problem, its MFs are similar to those obtained by TS-SPEA2$_{Acc^2}$ (even with any MFs that should be renamed).

5.4 Analysis on the Solution Selection and Importance of the Initialization

As said, an internal study on alternative possibilities to select solutions from final Pareto fronts and on the initialization influence is also performed by focusing on the electrical problem (the one with more possibilities to reduce not only the rule number but also the system error). Although we propose as final solution the most accurate one since our main objective is to reduce the number of rules but maintaining or improving the accuracy of the obtained models, there is another motivation that reinforced our decision. If this solution is maintained as a part of the final Pareto is because no other rule configuration is able to obtain a better parameter tuning (the main reason of the improved accuracy of these kinds of models). So we can be sure that this solution had the time to be evolved more or less in the proper way. However, the more we look for simpler solutions in the final Pareto the less we can be sure that these solutions had the time to be tuned (if a simpler solution appears at the end of the evolutionary process probably this solution had not the time to be properly tuned). In any case, the output of MOEAs really is a set of solutions from which an expert could choose the most convenient one.

In Table 5, we consider two different possibilities applied on the results obtained in Section 5.2 with TS-NSGA-II$_A$, TS-NSGA-II$_U$ and TS-SPEA2$_{Acc^2}$. In this case, we choose the solution with the best angle or utility measure in the Pareto fronts obtained from TS-NSGA-II$_A$ and TS-NSGA-II$_U$ respectively, and the i-th most accurate solution in the Pareto fronts obtained from

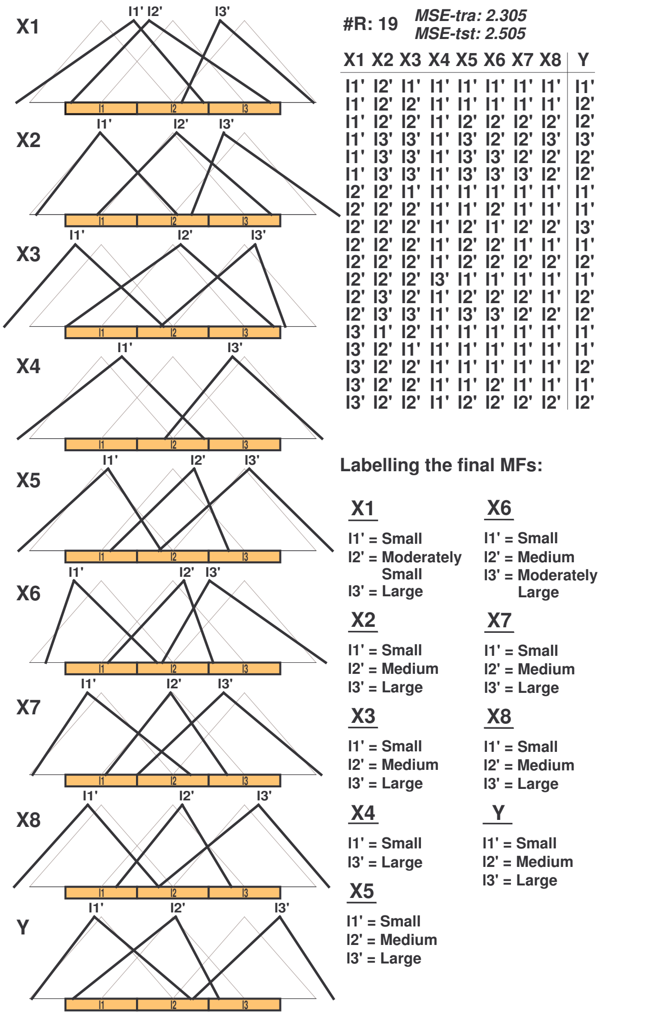


Fig. 8 DB with/without tuning (black/gray) and RB of a model obtained by TS-SPEA2$_{Acc^2}$ in the Abalon dataset

TS-SPEA2$_{Acc^2}$. As can be seen, the results obtained by choosing the proper i-th solution from TS-SPEA2$_{Acc^2}$ outperform the results obtained by knee based approaches. Since a knee is an ideal solution in the ideal Pareto but a promising one in a population of non dominated solutions, the knee based measure can be good to favor the evolution of promising solutions but it seems not so good to choose the final solution from evolved Pareto fronts that could present false knees (for example one solution with too few rules but without time to be properly tuned can be identified as a knee). In fact, standard deviations in the table show the diversity of solutions proposed considering knee measures (different knees appears in different runs, and for example a solution with 20 rules 27426 in training and 38432 in test is proposed by TS-NSGA-II$_U$-k in one of the 30 runs). In any case, independently of the mechanism considered to propose a final solution the most important thing is if the obtained front can be nearer of the optimal one, and the situation of the most accurate solution seems to be very indicative in this sense.

Table 5 Results obtained by choosing the knee or the i-th most accurate solution in the electrical distribution problem

Method	#R	MSE$_{tra}$	σ_{tra}	t	MSE$_{tst}$	σ_{tst}	t
TS-NSGA-II$_A$-k	25.5	13242	2383	+	17541	4184	+
TS-NSGA-II$_U$-k	24.2	15797	3945	+	20528	6802	+
TS-SPEA2$_{Acc^2}$-1	29.8	**10325**	**1121**	*	**13935**	**2759**	*
TS-SPEA2$_{Acc^2}$-2	28.3	10496	1126	=	14268	2925	=
TS-SPEA2$_{Acc^2}$-3	27.0	10835	1191	=	14460	2782	=
TS-SPEA2$_{Acc^2}$-4	25.9	11217	1307	+	14806	3069	=
TS-SPEA2$_{Acc^2}$-5	24.9	12194	2078	+	15417	3328	=

Table 6 Results obtained by random initialization of the C_S part (rule part) in the electrical distribution problem

Method	#R	MSE$_{tra}$	σ_{tra}	t	MSE$_{tst}$	σ_{tst}	t
TS-SPEA2	21.9	18768	2256	+	23951	**5198**	+
TS-NSGA-II	27.7	17688	2333	+	23762	7681	+
TS-NSGA-II$_A$	**19.5**	23981	3709	+	29442	7058	+
TS-NSGA-II$_U$	24.4	18728	2071	+	24148	5397	+
TS-SPEA2$_{Acc}$	23.2	**14175**	**1752**	*	**18289**	5571	*
TS-SPEA2$_{Acc^2}$	20.2	16539	2729	+	21977	5625	+
Best in Table 3:	29.8	10325	1121		13935	2759	

A study has been also performed on the importance of the initialization component for the rule selection part in the chromosome (considering a completely random initialization instead of the one presented in Section 4.1). Table 6 presents the results (again considering the most accurate solution in the final fronts) obtained by all the MOEAs considered for comparison starting with the initial rules selected at random (the best result in Table 3 is also included to show differences). By considering random initialization the results obtained present too low numbers of rules with much worse results especially in the test. This shows that to add rules that were not selected at the beginning is not easy since the MF parameters are quickly adapted to those rules that are just selected giving way to sub-optimal Pareto fronts. In any case, there are two important facts in these results:

– TS-SPEA2$_{Acc}$ and TS-SPEA2$_{Acc^2}$ methods were not very affected by the random initialization, presenting solutions that were also interesting from the trade-off point of view (very low number of rules and a good accuracy).
– Fixing the number of rules in the initial population can be a way to regulate the desired trade-off since this biases the number of rules in the final solutions.

6 Concluding Remarks

In this work, we have analyzed the application of different MOEAs to obtain simpler but still accurate linguistic fuzzy models by performing rule selection and a classic tuning of the MF parameters. In order to show the main differences with the previous works, a brief analysis of the state of the art on the use of MOEAs to get FRBSs with good accuracy-interpretability trade-off has been performed at first. From this study we can stress the following points:

- Most of the works only consider quantitative measures of the system complexity to determine the FRBS interpretability since the use of qualitative measures is still an open topic that needs of further and intense research efforts.
- None of the works (but the one in [5]) considered a learning or tuning of the MFs, only performing rule learning or selection.
- Algorithms considered were slight modifications of MOEAs proposed for general use (MOGA, NSGA-II, etc.) or specifically developed for this concrete and difficult problem. It is due to the special nature of this problem, in which to improve the accuracy objective is more difficult than simplifying the fuzzy models, by which the Pareto front finally obtained still becomes sub-optimal with respect to the accuracy objective. Therefore, MOEAs considering specific information about the problem are usually needed.

Since combining rule selection and tuning of the system parameters represents a more complex search space and therefore needs of different considerations with respect to the works in the existing literature, some considerations based on the experience are needed in the MOEA design process in order to get good solutions. From the results obtained, we can conclude that:

- The results obtained have shown that an appropriate use of MOEAs can represent a way to obtain even more accurate and simpler linguistic models than those obtained by only considering performance measures.
- Population initialization is an important component that can help to regulate the desired trade-off since this biases the number of rules in the final solutions. In this way, the results obtained by selecting all the rules in the initial population are able to find solutions in the most accurate Pareto zone.
- The best results were obtained by TS-SPEA2$_{Acc^2}$ on two different scenarios. These results show that the use of experience based knowledge in the MOEAs design process can significantly improve the search ability of these algorithms.

Finally, we would like to point out that the analysis presented in this work could help to extend this approach in order to consider other kinds of techniques or new interpretability measures for further works, e.g., another tuning types, learning, combination with the use of quality measures, etc.

References

1. 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 **10:9,** (2006) 717–734.
2. R. Alcalá, J. Alcalá-Fdez, F. Herrera and J. Otero, "Genetic learning of accurate and compact fuzzy rule based systems based on the 2-tuples linguistic representation," International Journal of Approximate Reasoning **44:1,** (2007) 45–64.
3. R. Alcalá, J. Alcalá-Fdez, M.J. Gacto and F. Herrera, "Rule base reduction and genetic tuning of fuzzy systems based on the linguistic 3-tuples representation," Soft Computing **11:5,** (2007) 401–419.
4. R. Alcalá, J. Alcalá-Fdez and F. Herrera, "A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection," IEEE Transactions on Fuzzy Systems **15:4,** (2007) 616–635.
5. R. Alcalá, M.J. Gacto, F. Herrera and J. Alcalá-Fdez, "A multi-objective genetic algorithm for tuning and rule selection to obtain accurate and compact linguistic fuzzy rule-based systems," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **15:5,** (2007) 539–557.
6. J. Branke, K. Deb, H. Dierolf, and M. Osswald, "Finding Knees in Multi-objective Optimization," Proc. Parallel Problem Solving from Nature Conf. - PPSN VIII, LNCS **3242,** (Birmingham, UK, 2004) 722–731.
7. J. Casillas, O. Cordón, F. Herrera, and L. Magdalena (Eds.), "Interpretability issues in fuzzy modeling," Studies in Fuzziness and Soft Computing **128,** (Springer, Heidelberg, Germany, 2003).
8. J. Casillas, O. Cordón, F. Herrera and L. Magdalena (Eds.), *Accuracy improvements in linguistic fuzzy modelling*, Studies in Fuzziness and Soft Computing, **129** (Springer, Heidelberg, Germany, 2003).
9. J. Casillas, O. Cordón, M.J. del Jesus, and F. Herrera, "Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction," IEEE Trans. Fuzzy Systems **13:1,** (2005) 13–29.
10. M. Cococcioni, P. Ducange, B. Lazzerini, F. Marcelloni, "A Pareto-based multi-objective evolutionary approach to the identification of mamdani fuzzy systems," Soft Computing **11,** (2007) 1013-1031.
11. C.A. Coello, and G. Toscano, "A Micro-Genetic Algorithm for multiobjective optimization," *First Int. Conf. on Evolutionary Multi-Criterion Optimization*, LNCS **1993,** (London, UK, 2001) 126–140.
12. C.A. Coello, D.A. Van Veldhuizen, and G.B. Lamont, "Evolutionary Algorithms for Solving Multi-Objective Problems," Kluwer Academic Publishers (2002).
13. O. Cordón, F. Herrera, and L. Sánchez, "Solving electrical distribution problems using hybrid evolutionary data analysis techniques," Applied Intelligence **10,** (1999) 5–24.
14. O. Cordon, F. Herrera, M.J. del Jesus, and P. Villar, "A multiobjective genetic algorithm for feature selection and granularity learning in fuzzy-rule based classification systems," Proc. IX IFSA World Congress and XX NAFIPS Int. Conf. (Vancouver, Canada, 2001) 1253–1258.
15. D. Corne, J. Knowles, and M. Oates, "The Pareto Envelope-based Selection Algorithm for multiobjective optimization," Proc. Parallel Problem Solving from Nature Conf. - PPSN VI, LNCS **1917,** (Paris, France, 2000) 839–848.
16. D. Corne, N. Jerram, J. Knowles, and M. Oates, "PESA-II: Region based selection in evolutionary multiobjective

optimization," Proc. Genetic and Evolutionary Computation Conf. (San Francisco, CA, 2001) 283–290.
17. K. Deb, "Multi-objective optimization using evolutionary algorithms," John Wiley & Sons (NY, USA, 2001).
18. K. Deb, S. Agrawal, A. Pratab, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II, " IEEE Trans. Evolutionary Computation **6:2,** (2002) 182–197.
19. M. Erickson, A. Mayer, and J. Horn, "The Niched Pareto Genetic Algorithm 2 applied to the design of groundwater remediation systems, " Proc. First Int. Conf. on Evolutionary Multi-Criterion Optimization, LNCS **1993,** (London, UK, 2001) 681–695.
20. L.J. Eshelman, "The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination," Foundations of genetic Algorithms **1,** (1991) 265–283.
21. L.J. Eshelman, and J.D. Schaffer, "Real-coded genetic algorithms and interval-schemata," Foundations of Genetic Algorithms **2,** (1993) 187–202.
22. C.M. Fonseca, and P.J. Fleming, "Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization," Proc. 5th Int. Conf. on Genetic Algorithms (San Mateo, CA, 1993) 416–423.
23. D. Goldberg, "A meditation on the computational intelligence and its future," Illigal Report #2000019, Department of General Engineering, University of Illinois at Urbana-Champaign, Foreword Proc. of the 2000 Int. Symposium on Computational Intelligence.
24. J. Horn, N. Nafpliotis, and D.E. Goldberg, "A niched Pareto genetic algorithm for multiobjective optimization," Proc. First IEEE Conf. on Evolutionary Computation, IEEE World Congress on Computational Intelligence **1,** (Piscataway, NJ, 1994) 82–87.
25. H. Ishibuchi, and T. Murata, "Multi-objective genetic local search algorithm," Proc. third IEEE Int. Conf. Evolutionary Computation (Japan, 1996) 119–124.
26. H. Ishibuchi, T. Murata, and I.B. Turksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," Fuzzy Sets and Systems **89:2,** (1997) 135–150.
27. H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule extraction," Information Sciences **136,** (2001) 109–133.
28. H. Ishibuchi, and T. Yamamoto, "Interpretability Issues in Fuzzy Genetics-Based Machine Learning for Linguistic Modelling, in: J. Lawry, J.G. Shanahan, A.L. Ralescu (Ed.), " Modelling with Words: Learning, Fusion, and Reasoning within a Formal Liguistic Representation Framework, LNCS **2873,** Springer-Verlag (Berlin, Heidelberg, 2003) 209–228.
29. H. Ishibuchi, and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," Fuzzy Sets and Systems **141:1,** (2004) 59–88.
30. H. Ishibuchi, and Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning," International Journal of Approximate Reasoning **44:1,** (2007) 4–31.
31. F. Jimenez, A.F. Gomez-Skarmeta, H. Roubos, and R. Babuska, "Accurate, Transparent, and Compact Fuzzy Models for Function Approximation and Dynamic Modeling through Multi-objective Evolutionary Optimization," Proc. First Int. Conf. on Evolutionary Multi-Criterion Optimization, LNCS **1993,** (Zurich, Switzerland, 2001) 653–667.
32. J.D. Knowles, and D.W. Corne, "Approximating the non dominated front using the Pareto archived evolution strategy," Evolutionary Computation **8:2,**(2000) 149–172.
33. K. Narukawa, Y. Nojima, and H. Ishibuchi, "Modification of evolutionary multiobjective optimization algorithms for multiobjective design of fuzzy rule-based classification systems," Proc. 2005 IEEE Int. Conf. on Fuzzy Systems (Reno, USA, 2005) 809–814.
34. E.H. Mamdani, and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," International Journal of Man-Machine Studies **7,** (1975) 1–13.
35. R.S. Rosenberg, "Simulation of genetic populations with biochemical properties." MA Thesis, Univ. Michigan, Ann Harbor, Michigan, (1967).
36. J.D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," Proc. First Int. Conf. on Genetic Algorithms (Pittsburgh, USA, 1985) 93–100.
37. N. Srinivas, and K. Deb, "Multiobjective optimization using nondominated sorting in genetic algorithms," Evolutionary Computation **2,**(1994) 221–248.
38. T. Takagi, and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," IEEE Trans. Systems, Man, and Cybernetics **15:1,** (1985) 116–132.
39. L.X. Wang, and J.M. Mendel, "Generating fuzzy rules by learning from examples," IEEE Trans. Systems, Man, and Cybernetics **22:6,** (1992) 1414–1427.
40. H.L. Wang, S. Kwong, Y.C. Jin, W. Wei, and K.F. Man, "Multi-Objective Hierarchical Genetic Algorithm for Interpretable Fuzzy Rule-Based Knowledge Extraction, " Fuzzy Sets and Systems **149:1,** (2005) 149–186.
41. H.L. Wang, S. Kwong, Y.C. Jin, W. Wei, and K.F. Man, "Agentbased evolutionary approach for interpretable rule-based knowledge extraction," IEEE Trans. Systems, Man, and Cybernetics Part C: Applications and Reviews **35:2,** (2005) 143–155.
42. S. Waugh, "Extending and Benchmarking Cascade-Correlation," PhD. thesis, Computer Science Department, University of Tasmania (1995).
43. E. Zitzler, and L. Thiele, "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach," IEEE Trans. Evolutionary Computation **3:4,** (1999) 257–271.
44. E. Zitzler, M. Laumanns, and L. Thiele, "SPEA2: Improving the strength Pareto evolutionary algorithm for multiobjective optimization," Proc. Evolutionary Methods for Design, Optimization and Control with App. to Industrial Problems (Barcelona, Spain, 2001) 95–100.