Research Article

Robustness Analysis of an Outranking Model Parameters' Elicitation Method in the Presence of Noisy Examples

Nelson Rangel-Valdez [iD],[1] **Eduardo Fernandez,**[2] **Laura Cruz-Reyes,**[1] **Claudia Gomez-Santillan** [iD],[1] **Gilberto Rivera,**[3] **and Rogelio Florencia** [iD][3]

[1]*Postgraduate and Research Division, National Mexican Institute of Technology/Madero Institute of Technology, 89440 Ciudad Madero, TAMPS, Mexico*
[2]*Faculty of Civil Engineering, Autonomous University of Sinaloa, 80040 Culiacan, SIN, Mexico*
[3]*National Lab of Information Technology, Autonomous University of Juarez City, 32310 Ciudad Juárez, CHIH, Mexico*

Correspondence should be addressed to Nelson Rangel-Valdez; nrangelva@conacyt.mx

One of the main concerns in Multicriteria Decision Aid (MCDA) is robustness analysis. Some of the most important approaches to model decision maker preferences are based on fuzzy outranking models whose parameters (e.g., weights and veto thresholds) must be elicited. The so-called preference-disaggregation analysis (PDA) has been successfully carried out by means of metaheuristics, but this kind of works lacks a robustness analysis. Based on the above, the present research studies the robustness of a PDA metaheuristic method to estimate model parameters of an outranking-based relational system of preferences. The method is considered robust if the solutions obtained in the presence of noise can maintain the same performance in predicting preference judgments in a new reference set. The research shows experimental evidence that the PDA method keeps the same performance in situations with up to 10% of noise level, making it robust.

1. Introduction

In Multicriteria Decision Aid (MCDA), one of its main concerns is the robustness of methods developed in this field. The term robust refers to the capacity for withstanding "vague approximations" and/or "zones of ignorance" to prevent the degradation of the properties that must be maintained [1]. Having this idea in mind, it is important to depict how robust a new method in MCDA is.

A wide variety of problems in decision aiding often involve multiple objectives to be minimized or maximized simultaneously. Because of the conflicting nature of the criteria, it is not possible to obtain a single optimum, and consequently, the ideal solution to a multiobjective optimization problem (MOP) cannot be reached. Therefore, the analysts resort to approaches that can handle multiple criteria and at the same time can shrink the number of solutions they provide to those concerning specific interests of a decision maker.

Several approaches that solve MOPs are based on Multiobjective Evolutionary Algorithms (MOEAs) and assume a model of the decision makers' (DM) preferences. This work focuses on the preference model proposed by Fernandez et al. in 2011. The model uses fuzzy outranking relations to incorporate preferences in MOEAs, such as the strict, weak, and K-preference. Also, it allows the mapping of a many-objective problem into a surrogate problem with only three objectives. The method has been applied in a wide variety of problems, including the portfolio problems with many objectives and project partial support [2].

In order to apply the preference model of Fernandez et al. [3], the outranking model's parameters must be elicited, for example, weights and thresholds required by the index of credibility of the outranking, a cutting level, and some additional symmetric and asymmetric parameters.

Information about the model's parameters can be obtained either directly or indirectly. On the one hand, the direct eliciting method has been criticized by Marchant [4]

and Pirlot [5] arguing that the only valid preference input information is that arising from the DM's preference judgments about actions or pairs of actions. As stated by Covantes et al. [6] and Doumpos et al. [7], these criticisms are even more significant in the frame of outranking methods, since the DM must set parameters that are very unfamiliar to her/him (e.g., veto thresholds). On the other hand, indirect elicitation methods use regression-inspired techniques for inferring the model's parameters from a set of decision examples [6, 7].

In the frame of outranking methods, preference-disaggregation analysis (PDA) approaches were pioneered by Mousseau and Slowinski [8]. They proposed to infer the ELECTRE TRI model's parameters (except veto thresholds) from a set of assignment examples by using nonlinear programming. Mousseau et al. [9] proposed a method to infer the weights from assignment examples through linear programming. Ngo The and Mousseau [10] used assignment examples to elicit the boundary profiles in ELECTRE TRI. Methods dealing with indirect elicitation of weights under inconsistent sets of assignment examples have been addressed by Mousseau et al. [11, 12]. Dias et al. [13] integrate interactively the elicitation phase with a robustness analysis.

Most of the related papers elude the inference of veto thresholds because eliciting all the parameters simultaneously requires solving a very complex nonlinear programming problem. Two papers proposed the use of evolutionary algorithms (EAs) to infer the entire set of ELECTRE model's parameters from a set of assignment examples (cf. [7, 14]). EAs are powerful tools for the treatment of nonlinearity and global optimization in polynomial time [15], as a more recent example, Álvarez et al. [16] used it to infer parameters that aid in the decision process at the collective level.

To the best of our knowledge, Fernandez et al. [17] was the first paper in which the reference information did not come from assignment examples, but from preference statements as *"x is at least as good as y"* and *"x is not at least as good as y."* This paper infers the entire set of the ELECTRE III model's parameters and the generalized outranking model with reinforced preferences proposed by Roy and Słowiński [18].

Cruz-Reyes et al. [19] proposed a PDA method to infer the entire parameter set of the relational system of preferences from Fernandez et al. [3]. This approach allows the introduction of the DM's preferential judgments through pairwise comparisons of different actions. However, this work lacks a robustness analysis which would allow measuring its capacity for withstanding vague approximations and/or zones of ignorance derived from its formal representation.

Based on the above, this research proposes a method for robustness analysis of the solutions offered by PDA methods based on metaheuristics. The study case is the Genetic Algorithm from the work of Cruz-Reyes et al. [19], which is used as a PDA method for the relational system of preferences proposed by Fernandez et al. [3]. The method is considered robust if it maintains the same performance with or without noise in the reference set; otherwise, it can be concluded that the method provides sensitive solutions. As a result, the experimental design proves the method robustness

by identifying that it estimates parameter value sets with a statistically nonsignificant difference when the noise levels are equal to or smaller than 10%.

Hence, the main contributions of this work are the proposed method for robustness analysis, and the noise model developed to introduce noise in a reference set. It is important to emphasize that both the method and model are the first in considering the full set of parameters in Fernandez et al. [3]. Also, an important part of the contributions of this work is the identification of the zone in the parameter space and the level of noise where the response sets are most compatible with the DM's preferences. It needs to be noted that the noise concept is related to the inconsistencies, or errors, between the preference model and the DM.

Aside from this introduction, the paper is organized as follows. Section 2 presents the optimization approach to estimate parameter values which are subject to the robustness analysis, the associated surrogate model, and the elements required for its definition. Section 3 shows the method followed in this work to perform the robustness analysis. Sections 4 and 5 present the experimental design conducted to evaluate the robustness of the optimization approach and the results obtained from it. Finally, Section 6 brings some concluding remarks derived from the research.

2. Optimization Approach for Inferring the Model's Parameter Values

This section is organized as follows. Firstly, it gives the definition of the optimization problem used as inference approach for the estimation of parameter values. This is followed by the optimization approach used to solve the studied problem and the description of the metaheuristic used. Finally, it presents the method that served as a basis to perform the analysis of robustness of the optimization approach for inference of the parameter values.

2.1. The Inference Approach. The best compromise is a solution of a problem associated with the DM's preferences. As was stated by Branke et al. [21], there has been an increasing interest in incorporating the DM's preference information in the search process. This situation is due to its influence on the reduction of the cognitive effort to identify a solution that best matches those preferences and to reinforce the selective pressure toward the Pareto frontier, for example, Cruz et al. [22].

A survey of strategies to incorporate preferences into multiobjective approaches can be found in [23, 24]. Particularly, this research deals with preference models based on outranking relations, such as the one developed by the works of Roy [20] and Fernandez et al. [3]. In these works, the preference model approaches situations concerning the behavior of real DMs using a relational system of preferences. The six binary relations that lie on that system are indifference, strict preference, weak preference, incomparability, K-preference, and nonpreference. These relations are associated with the predicate "the DM considers that option x is at least as good as y" through a degree of truth $\sigma(x, y)$ in [0, 1].

TABLE 1: Formal definition of outranking relations defined in [3, 20].

Relation	Notation	Formal Definition
(λ, β)-strictly preferred	$\mathbf{x}P(\eta, \lambda, \beta)\mathbf{y}$	(i) x dominates y; **OR** (ii) $\sigma(x, y) \geq \lambda \wedge \sigma(y, x) < 0.5$; **OR** (iii) $\sigma(x, y) \geq \lambda \wedge (0.5 \leq \sigma(y, x) < \lambda) \wedge (\sigma(x, y) - \sigma(y, x)) \geq \beta$
(λ, ε)-indifferent	$\mathbf{x}I(\eta, \lambda, \varepsilon)\mathbf{y}$	(i) $\sigma(x, y) \geq \lambda \wedge \sigma(y, x) \geq \lambda \wedge |\sigma(x, y) - \sigma(y, x)| < \varepsilon$
$(\lambda, \beta, \varepsilon)$-weakly preferred	$\mathbf{x}Q(\eta, \lambda, \beta, \varepsilon)\mathbf{y}$	(i) $\sigma(x, y) \geq \lambda \wedge \sigma(x, y) > \sigma(y, x)$; **AND** (ii) x not $P(\lambda, \beta)y$; **AND** (iii) x not $I(\lambda, \varepsilon)y$
(λ, ε)-K-preferred	$\mathbf{x}K(\eta, \lambda, \varepsilon)\mathbf{y}$	(i) $0.5 \leq \sigma(x, y) < \lambda$; **AND** (ii) $\sigma(y, x) < 0.5$; **AND** (iii) $(\sigma(x, y) - \sigma(y, x)) > \varepsilon$
Incomparability relation	$\mathbf{x}R(\eta)\mathbf{y}$	(i) $\sigma(x, y) < 0.5 \wedge \sigma(y, x) < 0.5$
$(\lambda, \beta, \varepsilon)$-nonpreference	$\mathbf{x}{\sim}(\eta, \lambda, \beta, \varepsilon)\mathbf{y}$	(i) not $xP(\lambda, \beta)y \wedge not$ $yP(\lambda, \beta)x$; **AND** (ii) not $xQ(\lambda, \beta, \varepsilon)y \wedge not$ $yQ(\lambda, \beta, \varepsilon)x$; **AND** (iii) not $xK(\lambda, \varepsilon)y \wedge not$ $yK(\lambda, \varepsilon)x$; **AND** (iv) not $xI(\lambda, \varepsilon)y$; **AND** (v) not xRy

Table 1 shows each outranking relation and its notation in columns one and two, and the necessary conditions that need to be satisfied for each relation in column three. The parameters η used in the computation of the function $\sigma(\eta, x, y)$ (cf. [19]), in combination with the credibility (λ), symmetry (β), and asymmetry (ε) thresholds, determine the preference relations.

Taking into account the fact that $\sigma(\eta, \mathbf{x}, \mathbf{y})$ is calculated in Table 1 as in ELECTRE III, then $\eta = (w, q, p, u, v)$ where (a) $w = \{w_1, w_2, \ldots, w_n\}$ is the vector of weights; (b) $q = \{q_1, q_2, \ldots, q_n\}$ is the vector of indifference thresholds; (c) $p = \{p_1, p_2, \ldots, p_n\}$ is the vector of preference thresholds; (d) $u = \{u_1, u_2, \ldots, u_n\}$ is the vector of veto thresholds; and (e) $v = \{v_1, v_2, \ldots, v_n\}$ is the vector of discordance thresholds, only when the simplification suggested by Mousseau and Dias [25] is used.

An inference approach for the previous relational system would be a strategy that could properly estimate the values of the parameters $(\eta, \lambda, \beta, \varepsilon)$; that is, it locates the configuration of the parameters that minimizes the inconsistencies among the DM's preferences and those identified by the model. This problem can be formally defined as follows. Let P_{fr} be the set of feasible parameter vectors $(\eta, \lambda, \beta, \varepsilon)$, $A = \{P, I, Q, K, \sim, R\}$ the set of preference relations, and T a reference set formed by pairs $(\mathbf{x}, \mathbf{y})$ of possible alternatives with an associate relation $\mathbf{x}A\mathbf{y}$ reflecting the preference judgment from a DM. The ideal solution of the parameter elicitation problem would be $(\eta_0, \lambda_0, \beta_0, \varepsilon_0) \in P_{\text{fr}}$ such that the following equivalences are satisfied for all $(\mathbf{x}, \mathbf{y}) \in T$:

$$\mathbf{O_1}: \mathbf{x}P\mathbf{y} \Longleftrightarrow \mathbf{x}P(\eta_0, \lambda_0, \beta_0)\mathbf{y}, \qquad (1a)$$

$$\mathbf{O_2}: \mathbf{x}I\mathbf{y} \Longleftrightarrow \mathbf{x}I(\eta_0, \lambda_0, \varepsilon_0)\mathbf{y}, \qquad (1b)$$

$$\mathbf{O_3}: \mathbf{x}Q\mathbf{y} \Longleftrightarrow \mathbf{x}Q(\eta_0, \lambda_0, \beta_0, \varepsilon_0)\mathbf{y}, \qquad (1c)$$

$$\mathbf{O_4}: \mathbf{x}K\mathbf{y} \Longleftrightarrow \mathbf{x}K(\eta_0, \lambda_0, \varepsilon_0)\mathbf{y}, \qquad (1d)$$

$$\mathbf{O_5}: \mathbf{x}{\sim}\mathbf{y} \Longleftrightarrow \mathbf{x}{\sim}(\eta_0, \lambda_0, \beta_0, \varepsilon_0)\mathbf{y}, \qquad (1e)$$

$$\mathbf{O_6}: \mathbf{x}R\mathbf{y} \Longleftrightarrow \mathbf{x}R(\eta_0)\mathbf{y}. \qquad (1f)$$

The best parameter setting should be the closest solution to the ideal one in the sense of certain acceptable metric. Following the works of Fernandez et al. [3, 17], we propose a metric based on the so-called *inconsistencies*. Shortly, an inconsistency arises when an equivalence in (1a)–(1f) is not fulfilled. For example, given a DM's judgment in $(\mathbf{x}, \mathbf{y}) \in T$ and inferred parameter values $(\eta, \lambda, \beta, \varepsilon) \in P_{\text{fr}}$, we call inconsistency with (1a) the fact that $\mathbf{x}$ is strictly preferred to $\mathbf{y}$ by the DM and however *not* $(\mathbf{x}P(\eta, \lambda, \beta)\mathbf{y})$, or vice versa, $\mathbf{x}P(\eta, \lambda, \beta)\mathbf{y}$ but $\mathbf{x}$ is not strictly preferred to $\mathbf{y}$ by the DM. Let $N_1, N_2, \ldots, N_6$ be the number of inconsistencies in T with (1a) through (1f), respectively.

Since the method by Fernandez et al. in 2011 gives priority to finding the nonstrictly outranked set, N_1 is by far the most important measure. The information coming from Q and K is also considered and turns out to be more important than that provided by I, R, and $\sim$. So, the best values for $(\eta, \lambda, \beta, \varepsilon)$ can be obtained from the solution of the following optimization problem:

$$\text{Minimize} \quad (E_1, E_2, E_3)$$
$$\text{s.t.} \quad (\eta, \lambda, \beta, \varepsilon) \in P_{\text{fr}}, \qquad (2)$$

where P_{fr} is the feasible region of parameter values, $E_1 = N_1$, $E_2 = N_3 + N_4$, and $E_3 = N_2 + N_5 + N_6$, with preemptive priority favoring E_1 over E_2 and E_2 over E_3. Note that, in this problem, xAy represents a preference judgment established by a DM, $\mathbf{x}A(\eta_0, \lambda_0, \beta_0, \varepsilon_0)\mathbf{y}$ represents an ideal preference statement derived from the outranking relational system, and $\mathbf{x}A(\eta, \lambda, \beta, \varepsilon)\mathbf{y}$ represents a preference statement from estimated parameters, which is derived from the outranking relational system. This problem has been studied by Cruz-Reyes et al. [19], and the PDA method used to properly
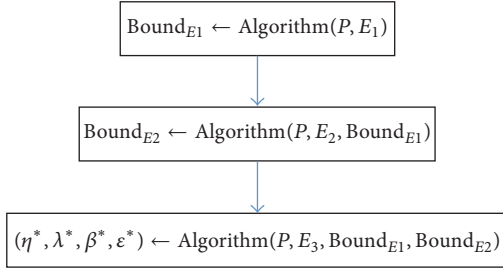
$$\text{Bound}_{E1} \leftarrow \text{Algorithm}(P, E_1)$$

$$\text{Bound}_{E2} \leftarrow \text{Algorithm}(P, E_2, \text{Bound}_{E1})$$

$$(\eta^*, \lambda^*, \beta^*, \varepsilon^*) \leftarrow \text{Algorithm}(P, E_3, \text{Bound}_{E1}, \text{Bound}_{E2})$$

FIGURE 1: Strategy followed to solve problem (2) using monoobjective algorithms.

Begin
   Step 1. $S_{\text{small}} \leftarrow \text{GenerateInstances}(\tau)$
   Step 2. $\tau^* \leftarrow \text{GeneticAlgorithm}(S_{\text{small}})$
   Step 3. $S'_{\text{small}} \leftarrow \text{NoiseModel}(S_{\text{small}})$
   Step 4. $\tau' \leftarrow \text{GeneticAlgorithm}(S'_{\text{small}})$
   Step 5. $S_{\text{large}} \leftarrow \text{GenerateInstances}(\tau)$
   Step 6a. $\Delta^*_{\text{perf}} \leftarrow \text{EvaluatePerformance}(\tau, \tau^*, S_{\text{large}})$
   Step 6b. $\Delta'_{\text{perf}} \leftarrow \text{EvaluatePerformance}(\tau, \tau', S_{\text{large}})$
   Step 7. $H_0 \leftarrow \text{RobustnessEvaluation}(\Delta^*_{\text{perf}}, \Delta'_{\text{perf}})$
End

ALGORITHM 1: Strategy followed to analyze the robustness of the PDA method.

estimate the values of its parameters is described in the following section.

2.2. Optimization Approach Based on a PDA Method. The multiobjective indirect elicitation method studied herein is the one proposed by Cruz-Reyes et al. [19]. The method is based on monoobjective optimization algorithms that solve problem (2). It exploits the preemptive priority established for the problem and finds the best value for the first objective; then, this value is used as a bound when the second objective is minimized. Finally, E_3 is minimized keeping the minimum values previously obtained for (E_1, E_2).

Figure 1 depicts the general procedure followed by the approach. Each metaheuristic algorithm solves $P = (\text{DM}_{\text{sim}}, T)$, that is, an instance of problem (2) formed by the parameter values that define the model of preferences for a simulated DM, denoted as DM_{sim}, and its reference set T, trying to minimize only the objective E_1. The best value Best_{E1} is used as a bound in the next step of the method, where the same algorithm seeks the minimization of objective E_2 while it validates that the solutions achieved a yield value on E_1 equal to or smaller than Best_{E1}. Finally, in the last step, the algorithm now minimizes the third objective of problem (2), that is, E_3, while keeping the values for E_1 and E_2 not greater than Best_{E1} and Best_{E2}, respectively. The best solution $(\eta^*, \lambda^*, \beta^*, \varepsilon^*)$ will arise from the parameter vectors obtained from the last step of this procedure.

The algorithm that was analyzed in this work was the Genetic Algorithm, the one with the best performance in the previous study. The evaluation function in each stage of the general procedure is related to objective E_i of problem (2). The encoding, the algorithms, and the fine-tuning process of the algorithms' parameters have been reported by Cruz-Reyes et al. [19].

3. Method for Robustness Analysis

Based on the work of [26], the concept of noise can be understood as random errors that are introduced in a sample. The presence of noise in an optimization process that seeks the adjustment of parameters of another strategy can affect the quality of its results. For example, a classification approach trained with an incorrect sample leads to misclassification of
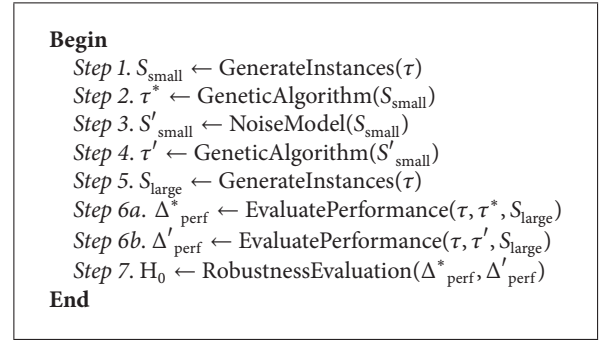
new outcomes, or a preference model with parameters estimated from a reference set containing erroneous preference judgments can derive inconsistencies with the DM's behavior. Because in many situations the presence of noise cannot be avoided, it is convenient to study the performance of methods that work over it and to identify how robust can they become in its presence.

This work uses the concept of noise to lead an investigation oriented to perform a robustness analysis of the strategy proposed by Cruz-Reyes et al. [19], summarized in the previous section. For this purpose, this section presents the general methodology to evaluate its quality, and details its components. The section is organized in three parts which involve (a) the presentation of the method; (b) the generation of random instances; (c) the model of noise for the instances. The definitions of the function that evaluates the quality of solutions of the optimization approach and the strategy followed to evaluate its robustness are given in the next section, related to the experimental design.

3.1. Description of the Method. Algorithm 1 depicts the method followed to analyze the robustness of the optimization approach described in Section 2. The method is developed in an ordered sequence of seven simple steps detailed in this section.

Firstly, in Step 1, the method randomly generates the set of instances S_{small}, which is formed by small reference sets T associated with a DM. In this work, the DM will be represented by a simulated DM and will be denoted by $\tau = (\eta, \lambda, \beta, \varepsilon)$. The generation of S_{small} is done through the *instance generator*.

In Step 2, after the construction of S_{small}, the method solves it by using the Genetic Algorithm proposed by Cruz-Reyes et al. in 2017, as *optimization approach* (i.e., the optimization approach based on a PDA method). The obtained solutions will be the best sets of parameter values τ^* that characterize the inference approach defined in problem (2) (see Section 2); these solutions represent estimations of parameter values of the outranking model for each instance in S_{small}.

Once the parameter values τ^* are computed, the method uses the *noise model* during Step 3 to introduce a small

amount of noise in S_{small} in order to produce a new set of instances S'_{small}. The noise is represented by a percentage of random modifications from correct preference statements toward wrong preference statements in S'_{small}. Then, in Step 4, the method, through the Genetic Algorithm, again infers the best sets of parameter values τ' of the outranking relations in the new set of instances S'_{small}.

The *performance evaluation* in the method is done in Steps 5 and 6. In Step 5, the method randomly generates a set of instances S_{large} formed by larger reference sets T. In Step 6, it compares the quality of the solutions in τ^* and τ' that were previously obtained by the Genetic Algorithm, against the correct parameter values τ, that were used to produce the instance sets S_{small} and S'_{small}. The indicators Δ^*_{perf}, Δ'_{perf} obtained in this step measure the number of inconsistencies found in S_{large} when using τ^* and τ', respectively, as inferred parameters of the preference model and τ as the real DM.

Finally, the *robustness analysis* of the Genetic Algorithm is analyzed statistically in Step 7. This step evaluates if there is a significant difference between the performance of the approach when using the noisy instances S'_{small} and its performance when it uses instances without errors in the preference relations, that is, S_{small}. The result on this step is the validity of the null hypothesis H_0.

In summary, the seven steps of the method involve five main components: (1) the instance generator; (2) the optimization approach; (3) the noise model; (4) the performance evaluation of the optimizer; and (5) the robustness analysis. These components are detailed in the remainder of this section with the exception of the Genetic Algorithm used as *optimization approach*, whose description can be seen in the work of Cruz-Reyes et al. in 2017.

3.2. Instance Generation.
Given that this work studies the robustness of an optimization approach that characterizes DM_{sim}, it requires sets of instances which can aid this purpose. Two types of sets were identified, the training set formed by small reference sets, denoted by S_{small}, and the testing set formed by large reference sets, denoted by S_{large}. A third set S'_{small} was created from S_{small} by adding some noise to it; as a result, both were used to estimate the parameter values of the preference model.

Each instance is characterized by (i) a reference set T formed by a finite number of alternatives and (ii) a set of statements made over pairs $(x, y) \in T \times T$, which are given by the DM as his/her preference judgments. The process of construction is depicted as follows.

A reference set T in any instance of S_i has i alternatives with 10 objectives whose values are randomly generated in the range $[1, 10]$. Every pair of alternatives $(x, y) \in T \times T$ has one preference statement defined by the DM. This statement is selected from $A = \{P, I, Q, K, R, \sim, P^{-1}, Q^{-1}, K^{-1}\}$, that is, the preferences that the real DM establishes in the instance.

This work defines $DM_{\text{sim-parameters}} = (\eta_0, \lambda_0, \beta_0, \varepsilon_0)$, or just τ_0, as a DM. The $DM_{\text{sim-parameters}}$ is constructed using random values for each criterion in the parameter vector $\eta_0 = \{w, q, p, u, v\}$ and one tuple $(\lambda_0, \beta_0, \varepsilon_0)$ from the set $\{(0.51, 0.15, 0.07), (0.67, 0.15, 0.07), (0.70, 0.20, 0.10),$

$(0.75, 0.20, 0.10)\}$. In order to simulate the DM's responses, we assume that (s)he states xA_iy if $A_i \in A$ and the relation $xA_i(\eta_0, \lambda_0, \beta_0, \varepsilon_0)y$ given in Table 1 results from τ_0. In other words, $xA_i(\eta_0, \lambda_0, \beta_0, \varepsilon_0)y$ is considered to be the response from the real DM when (s)he is questioned about his/her preference on a pair (x, y).

Particularly, the set of instances in this work are S_{10} and S_{20}, which use small reference sets with 10 and 20 alternatives, respectively. These small sets involve a manageable number of preference judgments for which a DM can still have time to define. The set of instances S_{100} uses large reference sets with 100 alternatives; this increased number of alternatives allows a better study of robustness of the inference approach.

The best set of solutions τ^* and τ' were contrasted in their ability to establish properly the preference relations. For this purpose, the 9900 preference judgments of each instance in S_{100} were used. The quality of τ^* and τ' was measured as the number of inconsistencies they had with respect to the real DM (the parameters τ), in the same preference statements. The method followed for generating these instances is described below.

3.3. Noise Model: Incorporation of Noise into Instances.
This section describes the noise model, that is, the general procedure used to insert noise in the instances generated through the instance generation method.

Given an instance of S_i, a level of noise is introduced on it through incorrect pairwise statements xAy, also denominated errors. For this purpose, the noise model uses three steps: (1) the specification of a percentage of error V; (2) the computation of the number of wrong statements; and (3) the modification of the statements in the instance.

In the first step, the model specifies the percentage V of preference statements that will be modified from the original reference set. These statements reflect the incorrect information present in the reference set T. Let us note that the percentage values V considered in this work were taken from $\{0\%, 5\%, 10\%, 20\%, 50\%\}$.

The second step of the model determines the amount of preference statements that are chosen to be incorrect. This value is denoted by $V\binom{i}{2}$, where $\binom{i}{2}$ is the maximum number of preferences defined with i alternatives, in the instance S_i, and V is the chosen percentage of such preferences that will be modified to represent errors.

Finally, in the third step, a total of $V\binom{i}{2}$ preference statements xAy are selected at random from T. Then, they are modified accordingly to its original statement under the opposite rules applied to the specific relation A involved, which are shown with (3a) through (3e) as conditionals.

$$xPy \longrightarrow xQy \vee xIy \vee xKy, \tag{3a}$$

$$xIy \longrightarrow xQy \vee x\sim y, \tag{3b}$$

$$xQy \longrightarrow xPy \vee xIy, \tag{3c}$$

$$xKy \longrightarrow xPy \vee xRy, \tag{3d}$$

$$xRy \longrightarrow x\sim y. \tag{3e}$$

The new set with the incorrect preference statements is denoted by S'_i. In the previous rules, the left-hand side denotes the original judgment xAy in $T \in S_i$, and the right-hand side represents the options for which they can be changed in S'_i (where $\vee$ is the disjunctive operator). Let us point out that whenever a preference statement xAy has more than one option, it is chosen at random with equal probability.

4. Experimental Design

The method to analyze the robustness of the optimization approach based on PDA is presented in Section 3. This section details the experiment conducted to implement such method, whose content is organized as follows. The first part describes the indicator of quality used to evaluate the solution set offered by the Genetic Algorithm; this indicator is based on the inconsistencies produced by the approach using the estimated value parameters obtained from each reference set T. The second part describes the robustness evaluation performed over the optimization approach; this part details the statistical analysis used to demonstrate whether there is significant difference between the performance of the optimization approach when using reference sets with and without noise.

4.1. Performance Evaluation. The method in Section 3.1 generates the sets $\tau^* = (\eta^*, \lambda^*, \beta^*, \varepsilon^*)$ and $\tau' = (\eta', \lambda', \beta', \varepsilon')$. These are sets of solutions produced by the PDA strategy that solves the sets of instances S_{small} and S'_{small} with and without noise, respectively. The quality of the sets τ^* and τ' is evaluated using the indicator I_P. This indicator measures the error in the prediction capacity of the sets as the differences in the estimation of preference statements. The details on this indicator are presented in the remainder of this section.

The calculation of the estimated error I_P requires $\tau_0 = (\eta_0, \lambda_0, \beta_0, \varepsilon_0)$, that is, the parameter value settings of $\text{DM}_{\text{sim-parameters}}$ that were used in the generation of a reference set T. The expected preference judgment $xA(\tau_0)y$ in T is compared for similarity against the estimated preference statements $xA(\tau')y$, and $xA(\tau^*)y$. Then, the error is measured as the amount of inconsistent preference statements that each solution set τ^* and τ' had accumulated with respect to the ones defined by τ_0.

Hence, the indicator I_P counts as an inconsistency each time that a strict preference $xP(\tau_0)y$ does not match the estimated relation $xA(\tau^*)y$ (or the relation $xA(\tau')y$ obtained from instances with noise). In other words, the parameter values obtained from PDA failed to predict the desired judgment and instead, produced a relation $A \neq P$. The indicator I_P is formally defined in (4).

Note that I_P can be seen as a measure of the level of discordance with the DM produced by the elicited parameters estimated by the PDA strategy. Also, let us point out that I_P is also known as the performance indicators Δ^*_{perf} and Δ'_{perf}; these symbols are used to represent the quality of the solutions from S_{small} and S'_{small}, respectively, in the robustness analysis method of Section 3.1. Finally, let us indicate that the sets of instances with small reference sets are

S_{10} and S_{20}, and the set of instances with large reference sets is S_{100}.

$$I_P = \sum_{(x,y) \in T \times T} f_1(x, y), \tag{4}$$

where $f_1(x, y)$ is a function whose value is 1 if the biconditional $xP(\tau_0)y \Leftrightarrow xP(\tau^*)y$ is false and 0 otherwise, and T is the reference set.

4.2. Robustness Evaluation. The experimentation on evolutionary algorithm implementations is needed to achieve better predictions about their performance and robustness. Statistical testing plays a central role to make the analysis of experiments on algorithms a more rigorous area. Many state-of-the-art publications report conclusions in terms of statistically meaningful coefficients such as p values [27, 28].

A *p value* is the probability, under an assumed model for the data including the null hypothesis that a statistical summary of the data (e.g., the sample mean of differences between two compared groups) would be equal to or more extreme than its observed value in the analyzed sample. The p value is an index of the incompatibility between the data in the sample and the proposed model. The smaller the p value, the greater the statistical incompatibility of the data with the null hypothesis. A small p value, for example, a value less than a level of significance of 0.05, offers some evidence against the null hypothesis. Likewise, a large p value gives some evidence in favor of the null hypothesis. A formal definition of p value is presented by Bartz-Beielstein et al. in 2010 [29].

The analysis of robustness of the Genetic Algorithm, used for parameter values estimation by Cruz-Reyes et al. in 2017, is performed by nonparametric statistical tests because the results of evolutionary algorithms regularly do not satisfy the assumptions of parametric tests [30]. The robustness evaluation is centered on the demonstration that there is no significant difference between the performance of Δ^*_{perf} and Δ'_{perf}. In other words, the quality of solutions measured by I_P has similar means when using a set of instances S_{small} or S'_{small} with or without noise.

The statistical analysis uses the indicators Δ^*_{perf} and Δ'_{perf} of the quality of solutions derived from the method in Section 3.1. The indicator Δ^*_{perf} is computed for each instance in the sets S_{10}, S_{20}. The value of the indicator Δ'_{perf} is also computed but using different levels of noise $V = \{5\%, 10\%, 20\%, 50\%\}$ for each instance in S'_{10} and S'_{20}. The different values for Δ^*_{perf} and Δ'_{perf} are grouped according to the level of noise (where a zero percent means no noise). After that, a statistical evaluation of their mean is carried out.

This work uses the STAC Web Platform (Statistical Web Tool found at URL: http://tec.citius.usc.es/stac/ranking .html), a statistical tool equipped with different statistical tests. Within this platform, the selected tests to perform the statistical evaluation were the *Non-parametric multiple groups One vs All*, with the particular case of *Friedman Aligned Ranks Test* as Ranking Test. Additionally, the following tests were specified as post hoc methods for p value adjustment: (a) Bonferroni-Dunn; (b) Holm; (c) Finner; and (d) Li.

During the statistical tests, the significance level was set at 0.05. In the Ranking Test, the null hypothesis H_0 was as follows: *The means of the results of two or more algorithms are the same*. In the post hoc analysis, the null hypothesis H_0 was as follows: *The mean of the results of the control method against other groups is equal (compared in pairs)*.

The results from the evaluation of the performance of each solution and from the analysis of robustness are given in the following section.

5. Results

This section presents the results derived from the analysis of robustness of the Genetic Algorithm that are organized in two subsections. The first subsection presents the summary of the values of performance indicator I_P from the estimations obtained with the sets of instances S_{10} and S_{20} to predict preference statements on the new set of instances S_{100}; in both cases, the results are shown in the presence and absence of noise. The second subsection summarizes the results from the statistical evaluation; according to this data, the method reveals tolerance of the optimization approach to the presence of noise of up to 10%, a result considered to be robust.

5.1. Results from Evaluation of the Performance. The first part of the experiment measured the performance using the indicator I_P. Table 2 presents its values achieved for each instance of the sets S_{10} and S_{20}. The first column of this table shows the number of instances, which were 40, or reference sets T, per set. From column two to six, the values of I_P are presented for S_{10} at $V = \{0\%, 5\%, 10\%, 20\%, 50\%\}$, that is, each different level of noise considered; observe that the value $V = 0\%$ corresponds to the case without noise. From column seven to eleven, the values of I_P are presented analogously for S_{20}.

From the results in Table 2, the accumulated number of inconsistencies varies by at most 250 from Δ^*_{perf}, with noise levels of up to 10%, in both sets. The difference is at least 500 inconsistencies when the noise level is 20% or greater. This result is similar when the instances are considered individually. For example, there are more than 20 instances in S_{10} and S_{20}, with noise level of up to 10% that has inconsistencies deviated at most 10% from that produced by Δ^*_{perf}; the number of such instances is considerably reduced for greater values of noise when keeping the same deviation value. This result raises the question, what level of noise can be considered different from a situation when it does not exist? This question is addressed in the next part of the experiment, where the use of statistical evaluation provided an answer to it. Note that, from these results, it can be considered that the level of disagreement increases, as it was expected, with the increment in the level of noise.

5.2. Results from the Statistical Evaluation. Table 3 summarizes the results obtained from *Ranking* through the *Friedman Aligned Ranks Test*. It presents the *p value* obtained for the considered relations, as well as the result of the hypothesis H_0 that is being validated. Note that there is significant difference for the relations *xPy*.

Tables 4(a) and 4(b) summarize the results of post hoc analysis, for S_{10} and S_{20}, respectively. The relationships where the null hypothesis H_0 is accepted (i.e., that there is NO significant difference in the performance of the PDA method) are highlighted in bold. The results clearly show that the comparisons between the reference set without noise (noise level of 0%) against the other strategies present significant difference from a noise level of 20%. These results support the claim that the significant difference in the quality of the results of the PDA method studied starts with a noise level of 20%.

According to the results obtained, the method of learning the preferences of a DM based on the PDA method proposed by Cruz-Reyes et al. [19] proves robust, since it is statistically demonstrated that there is no significant difference in the quality of solutions provided by the method, when applied to two different sets of instances S_{10} and S_{20}, each formed of reference sets with 10 and 20 alternatives, respectively. The conducted experiment showed that with noise levels V of up to 10%, the solutions produced by the method had the same performance in predicting the preference statements as a new set S_{100} formed by instances with larger reference sets T.

On the other hand, the method starts to have a significant difference in performance when the error rate is greater than 10%. Therefore, when the DM introduces a small number of errors in his/her preferences within the provided reference set, the analyzed optimization approach, based on the PDA method, may still produces an appropriate estimate of parameter values of the ad hoc preference model used by the DM. These results make the method robust (i.e., fault tolerant).

6. Concluding Remarks

The present work proposes a method for robustness analysis of PDA strategies that work with the complete set of the ELECTRE III model's parameters and the generalized outranking model with reinforced preferences proposed by Roy and Słowiński [18]. The method does the parameters' elicitation with different levels of noise and validates whether there is or not significant difference in the use of such values to model preferences of a decision maker (DM).

Another contribution of this research is the noise model. This model arises from the need, in the robustness analysis, of defining bias in the reference set toward a region that represents incorrect preference statements of a DM, that is, his/her errors. The proposed noise model randomly replaces a percentage of the original preference statements by others that are closely related (resembling a vague approximation of the real desires of a DM).

The main finding derived from the experiments on the analyzed case of study, the PDA method of Cruz-Reyes et al. [19], was the identification that the method elicits parameters with a tolerance of up to 10% of error in the reference set. This tolerance means that even though the DM could introduce by mistake some wrong preference statements in the reference set, the PDA method is still able to estimate proper parameter values for the preference model that characterize him/her. Within this context, it could be possible to use the proposed

TABLE 2: Results from the evaluation of the performance of the optimization approach.

Instance	I_P = inconsistencies on O_1 (xPy) with set S_{10}					I_P = inconsistencies on O_1 (xPy) with set S_{20}				
	Noise level									
	Δ^*_{perf}	Δ'_{perf}	Δ'_{per}	Δ'_{per}	Δ'_{per}	Δ^*_{perf}	Δ'_{perf}	Δ'_{per}	Δ'_{per}	Δ'_{per}
	0%	5%	10%	20%	50%	0%	5%	10%	20%	50%
1	60	68	84	65	82	65	65	64	79	61
2	33	43	52	67	32	100	134	104	132	109
3	21	22	29	21	96	41	61	90	80	94
4	24	19	18	77	120	37	57	87	82	35
5	75	106	70	56	135	89	93	76	71	134
6	107	181	35	111	117	102	82	64	120	146
7	64	43	42	202	196	66	75	87	124	112
8	99	80	82	94	33	68	39	74	53	57
9	68	64	79	76	80	40	57	55	46	54
10	77	116	78	124	148	76	55	59	114	89
11	71	51	49	24	149	31	42	35	33	115
12	31	111	31	172	97	21	24	50	35	80
13	123	117	61	97	65	39	30	36	138	48
14	64	36	57	41	132	58	43	93	83	85
15	59	48	62	54	43	27	31	32	63	54
16	38	50	55	178	184	27	56	40	37	50
17	69	65	75	142	103	43	52	37	62	169
18	53	57	161	87	41	35	127	49	41	137
19	53	78	67	112	79	32	24	31	85	168
20	69	69	71	57	140	68	43	40	91	41
21	122	65	109	73	95	26	44	30	31	69
22	54	146	61	86	77	68	46	40	159	50
23	40	59	33	173	180	66	36	38	181	154
24	79	71	133	271	62	31	25	26	51	63
25	65	69	74	89	68	36	42	76	62	153
26	45	60	34	43	53	114	63	116	152	122
27	39	15	117	34	125	62	40	37	72	62
28	74	14	32	17	31	72	68	104	25	32
29	64	56	80	63	91	37	32	86	45	149
30	144	113	44	48	67	23	31	22	27	30
31	110	126	99	116	106	113	145	142	87	138
32	32	49	28	103	83	31	44	31	36	64
33	62	82	46	76	100	22	34	55	15	68
34	78	45	54	34	110	37	36	82	25	39
35	46	76	41	62	53	44	54	44	53	76
36	54	41	46	175	25	31	43	29	38	53
37	77	93	83	87	70	43	37	50	52	139
38	32	44	45	56	51	41	49	53	56	35
39	25	26	28	121	97	92	88	84	29	93
40	160	140	117	142	180	72	61	22	21	76

TABLE 3: Summary of results from Ranking Test using *Friedman Aligned Ranks*.

Set	Relation	Adjusted p value	Result
S_{10}	xPy	0.00023	H_0 rejected
S_{20}	xPy	0.00000	H_0 rejected

TABLE 4: Results of post hoc analysis for sets S_{10} and S_{20}. Summary of the adjusted p values obtained from different tests.

(a) Adjusted p values obtained for set S_{10} with different tests

Test	Indicator	Comparisons among levels of noise			
		0% versus 5%	0% versus 10%	0% versus 20%	0% versus 50%
Bonferroni-Dunn	I_P	**1.00000**	**1.00000**	0.00954	0.00121
Holm	I_P	**1.00000**	**1.00000**	0.00716	0.00121
Finner	I_P	**0.60988**	**0.76832**	0.00476	0.00120
Li	I_P	**0.68609**	**0.76832**	0.01019	0.00130

(b) Adjusted p values obtained for set S_{20} with different tests

Test	Indicator	Comparisons among levels of noise			
		0% versus 5%	0% versus 10%	0% versus 20%	0% versus 50%
Bonferroni-Dunn	I_P	**1.00000**	**0.73796**	0.02730	0.00000
Holm	I_P	**0.55575**	**0.36898**	0.02407	0.00000
Finner	I_P	**0.55575**	**0.23809**	0.01360	0.00000
Li	I_P	**0.55575**	**0.29343**	0.01513	0.00000

method on other PDA strategies in the frame of outranking approaches that handle reference set as pairs of alternatives.

It is important to note that this work assumes that the reference set is provided as preference statements over pairs of alternatives and does not consider cases where the preference information could be given by a set of sorted alternatives. For the latter case at least a different noise model would be required, and hence, it would constitute a line of research for future work.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] H. Aissi and B. Roy, "Robustness in multi-criteria decision aiding," in *Trends in Multiple Criteria Decision Analysis*, M. Ehrgott, Ed., vol. 142 of *International Series in Operations Research & Management Science*, pp. 87–121, Springer, New York, NY, USA, 2010.

[2] E. Fernandez, C. Gomez, G. Rivera, and L. Cruz-Reyes, "Hybrid metaheuristic approach for handling many objectives and decisions on partial support in project portfolio optimisation," *Information Sciences*, vol. 315, pp. 102–122, 2015.

[3] E. Fernandez, E. Lopez, F. Lopez, and C. A. Coello Coello, "Increasing selective pressure towards the best compromise in evolutionary multiobjective optimization: the extended NOSGA method," *Information Sciences*, vol. 181, no. 1, pp. 44–56, 2011.

[4] T. Marchant, *How to assign numerical values to parameters aiming at differentiating the role of criteria in a comprehensive preference model? Round Table debate at the 71st Meeting of the Euro Working Group on Multiple Criteria Decision Aiding*, Italy, Torino, 2010.

[5] M. Pirlot, "Weights of criteria [in French]," in *Proceedings of the Round Table debate at the 71st Meeting of the Euro Working Group on Multiple Criteria Decision Aiding*, Torino, Italy, 25-27 March.

[6] E. Covantes, E. Fernandez, and J. Navarro, "Handling the multiplicity of solutions in a MOEA-based PDA-THESEUS framework for multi-criteria sorting," *Foundations of Computing and Decision Sciences*, vol. 41, no. 4, pp. 213–235, 2016.

[7] M. Doumpos, Y. Marinakis, M. Marinaki, and C. Zopounidis, "An evolutionary approach to construction of outranking models for multicriteria classification: The case of the ELECTRE TRI method," *European Journal of Operational Research*, vol. 199, no. 2, pp. 496–505, 2009.

[8] V. Mousseau and R. Slowinski, "Inferring an ELECTRE TRI Model from Assignment Examples," *Journal of Global Optimization*, vol. 12, no. 2, pp. 157–174, 1998.

[9] V. Mousseau, J. Figueira, and J.-P. Naux, "Using assignment examples to infer weights for ELECTRE TRI method: some experimental results," *European Journal of Operational Research*, vol. 130, no. 2, pp. 263–275, 2001.

[10] A. Ngo The and V. Mousseau, "Using assignment examples to infer category limits for the ELECTRE TRI method," *Journal of Multi-Criteria Decision Analysis*, vol. 11, no. 1, pp. 29–43, 2002.

[11] V. Mousseau, J. Figueira, L. C. Dias, C. G. da Silva, and J. Climaco, "Resolving inconsistencies among constraints on the parameters of an MCDA model," *European Journal of Operational Research*, vol. 147, no. 1, pp. 72–93, 2003.

[12] V. Mousseau, L. C. Dias, and J. Figueira, "Dealing with inconsistent judgments in multiple criteria sorting models," *4OR*, vol. 4, no. 2, pp. 145–158, 2006.

[13] L. Dias, V. Mousseau, J. Figueira, and J. Clímaco, "An aggregation/disaggregation approach to obtain robust conclusions with ELECTRE TRI," *European Journal of Operational Research*, vol. 138, no. 2, pp. 332–348, 2002.

[14] E. Fernandez, J. Navarro, and S. Bernal, "Multicriteria sorting using a valued indifference relation under a preference disaggregation paradigm," *European Journal of Operational Research*, vol. 198, no. 2, pp. 602–609, 2009.

[15] C. A. Coello, G. B. Lamont, and D. A. Van Veldhuizen, *Algorithms for SolvingMulti_Objective Problems, Second Edition*, Springer, New York, NY, USA, 2nd edition, 2007.

[16] P. A. Álvarez, D. C. Morais, J. C. Leyva, and A. T. Almeida, "A multi-objective genetic algorithm for inferring inter-criteria parameters for water supply consensus," in *Evolutionary Multi-Criterion Optimization*, A. Gaspar-Cunha, C. Henggeler Antunes, and C. Coello, Eds., vol. 9019 of *Lecture Notes in Computer Science*, pp. 218–233, Springer, New York, NY, USA, 2015.

[17] E. Fernandez, J. Navarro, and G. Mazcorro, "Evolutionary multi-objective optimization for inferring outranking model's parameters under scarce reference information and effects of reinforced preference," *Foundations of Computing and Decision Sciences*, vol. 37, no. 3, pp. 163–197, 2012.

[18] B. Roy and R. Słowiński, "Handling effects of reinforced preference and counter-veto in credibility of outranking," *European Journal of Operational Research*, vol. 188, no. 1, pp. 185–190, 2008.

[19] L. Cruz-Reyes, E. Fernandez, and N. Rangel-Valdez, "A metaheuristic optimization-based indirect elicitation of preference parameters for solving many-objective problems," *International Journal of Computational Intelligence Systems*, vol. 10, pp. 56–77, 2017.

[20] B. Roy, *Multicriteria Methodology for Decision Aiding*, Kluwer, Alphen aan den Rijn, Netherlands, 1996.

[21] J. Branke, S. Corrente, S. Greco, R. Slowiński, and P. Zielniewicz, "Using Choquet integral as preference model in interactive evolutionary multiobjective optimization," *European Journal of Operational Research*, vol. 250, no. 3, pp. 884–901, 2016.

[22] L. Cruz, E. Fernandez, C. Gomez, G. Rivera, and F. Perez, "Many-objective portfolio optimization of interdependent projects with a priori incorporation of decision maker preferences," *Applied Mathematics & Information Sciences*, vol. 8, no. 4, pp. 1517–1531, 2014.

[23] S. Bechikh, *Incorporating Decision Makers Preference Information in Evolutionary Multi-objective Optimization [Ph.D. thesis]*, High Institute of Management of Tunis, University of Tunis, Tunisia, 2013, http://delta.cs.cinvestav.mx/~ccoello/EMOO/thesis-bechikh.pdf.gz.

[24] J. Branke, "MCDA and multiobjective evolutionary algorithms," in *Multiple Criteria Decision Analysis: State of The Art Surveys*, S. Greco, M. Ehrgott, and J. Figueira, Eds., pp. 10–1007, Springer, New York, NY. USA, 2015.

[25] V. Mousseau and L. Dias, "Valued outranking relations in ELECTRE providing manageable disaggregation procedures," *European Journal of Operational Research*, vol. 156, no. 2, pp. 467–482, 2004.

[26] D. Angluin and P. Laird, "Learning From Noisy Examples," *Machine Learning*, vol. 2, no. 4, pp. 343–370, 1988.

[27] S. García, D. Molina, M. Lozano, and F. Herrera, "A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 Special Session on Real Parameter Optimization," *Journal of Heuristics*, vol. 15, no. 6, pp. 617–644, 2009.

[28] N. Veček, M. Črepinšek, and M. Mernik, "On the influence of the number of algorithms, problems, and independent runs in the comparison of evolutionary algorithms," *Applied Soft Computing*, vol. 54, pp. 23–45, 2017.

[29] T. Bartz-Beielstein, M. Chiarandini, L. Paquete, and M. Preuss, Eds., *Experimental Methods for the Analysis of Optimization Algorithms*, Springer, New York, NY, USA, 2010.

[30] S. Garcia, A. Fernandez, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.