A Statistical Analysis of Performance in the 2021 CEC-GECCO-PESGM Competition on Evolutionary Computation in the Energy Domain

Fernando Lezama, Joao Soares, and Bruno Canizes
GECAD, Polytechnic of Porto,
Porto, Portugal
Email: {flz, jan, bmc}@isep.ipp.pt

Zita Vale
Polytechnic of Porto,
Porto, Portugal
Email: zav@isep.ipp.pt

***Abstract*—Evolutionary algorithms (EAs) have emerged as an efficient alternative to deal with real-world applications with high complexity. However, due to the stochastic nature of the results obtained using EAs, the design of benchmarks and competitions where such approaches can be evaluated and compared is attracting attention in the field. In the energy domain, the "2021 CEC-GECCO-PESGM Competition on Evolutionary Computation in the Energy Domain: Smart Grid Applications" provides a platform to test and compare new EAs to solve complex problems in the field. However, the metric used to rank the algorithms is based solely on the mean fitness value (related to the objective function value only), which does not give statistical significance to the performance of the algorithms. Thus, this paper presents a statistical analysis using the Wilcoxon pair-wise comparison to study the performance of algorithms with statistical grounds. Results suggest that, for track 1 of the competition, only the winner approach (first place) is significantly different and superior to the other algorithms; in contrast, the second place is already statistically comparable to some other contestants. For track 2, all the winner approaches (first, second, and third) are statistically different from each other and the rest of the contestants. This type of analysis is important to have a deeper understanding of the stochastic performance of algorithms.**

***Index Terms*—Evolutionary computation, metaheuristics, power systems, optimization, smart grids**

I. Introduction

The evolution of the electrical grid adopting new technologies has given place to smart grids (SG), intelligent networks that promise different stakeholders in power systems [1]. This shift in the paradigm of the energy field, together with the high penetration of Distributed Generation (DG), poses a new level of complexity in the planning, management, and operation of power and energy systems. Therefore, utilities, governments, and R&D centers are trying to find ways to cope with the challenges that such a complex and dynamic environment brings [2]. For instance, the consideration of uncertainty associated with stochastic renewable generation turns the mathematical formulations of some optimization problems almost intractable without a huge deal of assumptions and simplifications, making the solutions unrealistic in real-world scenarios [3].

One of the alternatives to find solutions to complex problems in the energy domain, while attracting the interest of research centers in solving such problems, relays in the organization of worldwide competitions launched at major events and conferences [4], [5]. This article presents the two tracks launched at the 2021 CEC-GECCO-PESGM Competition on Evolutionary Computation in the Energy Domain: Smart Grid Applications, an initiative that has been running already for several years and is becoming a valuable reference to test and compare state-of-the-art algorithms using computational intelligence (CI).

Track 1 "Bi-level optimization of bidding strategies in local energy markets (LEM)" follows the same model as the 2020 competition edition in which a bi-level problem for bidding in local energy markets is formulated [6]. In this competitive environment, agents search for profits at the upper level (multi-leader problem), while their energy transactions are maximized at the lower level (single-follower problem). The bids/offers of agents in the upper level set the clearing price in the lower level resulting in a strong interdependence of their decisions.

Track 2 "Flexibility management of home appliances to support DSO requests", presents a mixed-integer non-linear programming (MINLP) model. The model represents an aggregator that provides flexibility to a distribution system operator (DSO) using the load flexibility coming from end-users. The aggregator and end-users are assumed to be equipped with technology and infrastructure that enable the re-schedule of shifting/real-time home-appliances to provide a request from a DSO. In addition, remuneration to the households participating in the demand response program according to their preferences and the modification of their baseline profile is also considered [7].

This article presents briefly how the 2021 "Competition on Evolutionary Computation in the Energy Domain: Smart

This research has received funding from FEDER funds through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-028983; by National Funds through the FCT Portuguese Foundation for Science and Technology, under Projects PTDC/EEI-EEE/28983/2017(CENERGETIC),CEECIND/02814/2017, and UIDB/000760/2020.

Grid Applications" was organized, explaining the competition framework briefly. We also provide the datasets and results of the competitors and provide a Wilcoxon statistical test of the top three winner approaches. 16 algorithms submit entries to track 1, and 16 algorithms to Track 2 (a total of 20 different algorithms). The algorithms were developed, tuned, and tested by different researchers around the world. A predefined computing budget of function evaluations was established for each track, and the performance of each algorithm was compared considering the fitness value over a given number of runs. While evaluating the performance of algorithms based on the average fitness is more practical, we cannot claim that an algorithm is statistically superior (or even different) to rest based on that metric [8]. Therefore, statistical analysis is used in this paper to investigate the performance of the winner approaches in more detail.

The article is organized as follows. Section II presents the competition schedule and present the two proposed tracks. Section III presents the simulation framework and introduces the algorithms submitted to the competition. Section IV provides the basis of the statistical pair-wise comparison Wilcoxon test used to evaluate the performance of the winner approaches. Section V contrast the results of the evaluation criteria used in the competition and the results obtained with the statistical test. Finally, Section VI concludes the paper and provides some final discussion on the findings.

II. COMPETITION SCHEDULE AND TRACKS

The 2021 Competition on "Evolutionary Computation in the Energy Domain: Smart Grid Applications" has been launched at three major events, the IEE Congress on Evolutionary Computation (CEC), the ACM Genetic and Evolutionary Computation Conference (GECCO), and the IEEE Power and Energy Society General Meeting (PESGM), to bring together and test state of the art algorithms applied to challenging problems in the energy domain. These alternative methods are attracting the attention of practitioners due to their potential of dealing with complexities in some mathematical problems such as high-dimensionality, non-linearity, non-convexity, multimodality, or discontinuity in the search space [3]. Furthermore, understanding the validity of the "no-free lunch theorem" [9], we provide a coherent simulation framework where participants can test CI algorithms solving real-world applications beyond typical benchmarks and standardized CI problems.

The 2021 edition of the competition, supported by computational intelligence society (CIS), the working group on modern heuristic optimization (WGMHO), and the Intelligent System Application Technical Committee (ISATC) task force 3 Computational Intelligence in the Energy Domain, introduces two independent tracks:

Track 1: Bi-level optimization of bidding strategies in local energy markets Track 2: Flexibility management of home appliances to support DSO requests

The guidelines and rules, as well as the schedule of the conference can be found at "http://www.gecad.isep.ipp.pt/ERM-competitions/2021-2/". The simulation framework and algorithms that took part in the competition can be found there; thus, the tracks are an open challenge to CI practitioners interested in solving such problems. The platform was implemented and tested in MATLAB©. The schedule of the competition and the major events in which the results were considered and presented was:

- 15 January 2021: Call for competition.
- February 21, 2021: Submission of articles to CEC SS-44 Evolutionary Algorithms for Complex Optimization in the Energy Domain
- April 12, 2021: Submission of 2-page papers to be included in the GECCO Companion.
- 15 June 2021: Submission of results and codes.
- 28 June - 01 July, 2021: Announcement of the best three ranked algorithms at CEC 2021.
- 10 July - 14 July, 2021: Announcement of the best three ranked algorithms at GECCO 2021.
- 28 July, 2021: Presentation of the winners at the IEEE PES General Meeting.

In the following subsections, the two tracks are described briefly.

A. Track 1: Bi-level optimization of bidding strategies in local energy markets

In track 1, a bilevel optimization problem models a bidding procedure in local energy markets (LEM). Figure 1 shows the simulated environment in this track, in which players (also called agents) submit bids to minimize their costs and offers to maximize their profits. Three types of agents are considered: consumers, small producers, and prosumers (i.e., consumers with PV generation). The LEM (lower level) responds to the bids and offers of agents (upper level), maximizing the energy transacted and sending the clearing price cp_t and corresponding energy transactions $X_{i,j}$ to all the market participants. It is assumed that an aggregator/retailer acts as a backup to guarantee demand supply and avoid balance deviations due to PV and load uncertainty.

Considering a set of consumer agents $i = \{1, 2, ..., N_c\}$, and producer agents $j = \{1, 2, ..., N_p\}$, the upper level (multi-leader problem) captures the minimization of costs of each agent i (Eq. 1), and the maximization of profits of each agent j (Eq. 2):

$$\text{minimize} \quad C_i = \sum_{t=1}^{T} \left(\sum_{j} cp_t * x_{j,i,t} + c_t^{\text{agg}} * E_{i,t}^{\text{buy}} \right) \tag{1}$$

$$\text{maximize} \quad P_j = \sum_{t=1}^{T} \left(\sum_{i} cp_t * x_{j,i,t} + c_t^{\text{F}} * E_{j,t}^{\text{sell}} - c_t^{\text{m}} * G_{j,t} \right) \tag{2}$$

where cp_t is the LEM clearing price (equal for buyers and sellers); $x_{j,i,t}$ contains the energy sold/bought by agent i/j to agent j/i in the LEM; c_t^{agg} is the aggregator tariff and $E_{i,t}^{\text{buy}}$ is the energy buy by agent i from the grid; c_t^{F} is the feed-in tariff and $E_{j,t}^{\text{sell}}$ is the energy sold by agent j to the

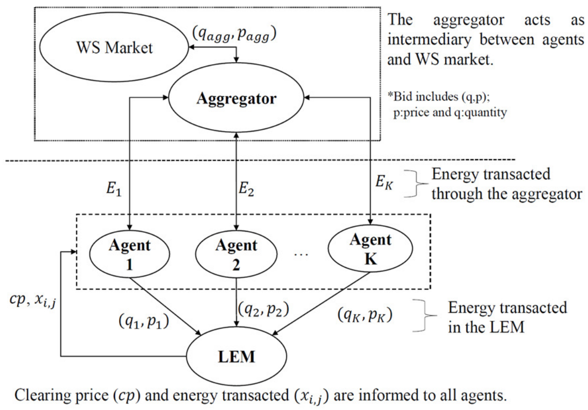

Fig. 1. Track 1 considers a local market with different types of agents.

grid; $c_t^m * G_{j,t}$ is the marginal cost associated to producer j. The formulation is also subject to constraints such as the energy balance constraints and supply demand constraints (not included here due to space limitations).

On the other hand, the lower level problem (single-follower problem) maximizes the energy transacted in LEM according to:

$$\text{maximize} \quad X^{\text{LEM}} = GE_v \cup DE_w \tag{3a}$$
$$st.$$
$$cp_t = \max(s_j(GE_v)) \tag{3b}$$
$$cp_t, X^{\text{LEM}} \geq 0, \tag{3c}$$

where X^{LEM} is the energy transacted in the LEM; GE is a set containing the offers of energy g_j in ascending order of price; DE is a set containing the bids for energy d_i in descending order of price; GE_v and DE_w describe the aggregated amount of bids and offers that comply with $s_v \leq s_w$ (i.e., when the supply and demand curves intersect). Thus, the clearing price at each time t (cp_t) is determined by the highest offer still accepted in the LEM merit order mechanism. Notice that the cp is determined in the lower level by applying a merit-order procedure, and therefore, depends directly on the decisions taken at the upper level. Also, as explained in [], we assume that agents trade energy in the LEM with prices in the range of a feed-in tariff cf and a wholesale market plus an aggregator fee $cAGG$. With this assumption, and assuming that $cF < cAGG$, we guarantee that buying or selling energy to the aggregator is less beneficial than participating in the LEM.

More details about the formulation of the problem is available in the publication [6].

B. Track 2: Flexibility management of home appliances to support DSO requests

In track 2, mixed-integer non-linear programming (MINLP) formulation models the flexibility management of home appliances by an aggregator to support DSO requests. Figure 2 shows the simulation environment of track 2, in which an aggregator is in control of the management of devices with demand response capabilities. Furthermore, considering that users register voluntarily for participation in flexibility provision, receiving monetary compensations for it, the aggregator is enabled to modify their baseline profiles by shifting or regulating the power of some home appliances.

The objective function in this problem is modeled as in [7]:

$$\text{Minimize } f = \left(\sum_{i=1}^{N_I} Rem_{A(i)} + \sum_{j=1}^{N_J} Rem_{B(j)} \right) + C_{DSO} \cdot F_{match} \tag{4}$$

$$Rem_{A(i)} = \begin{cases} C_{A(i)} & if \quad t_{start(i)} \neq t_{new(i)} \\ 0 & \text{otherwise} \end{cases}$$

$$Rem_{B(j)} = C_{B(i)} \cdot \sum_{t=1}^{N_T} \left| B_{base(j,t)} - B_{flex(j,t)} \right|$$

$$F_{match} = \sum_{t=1}^{N_T} \left| F_{agg(t)} - F_{DSO(t)} \right|$$

where the first term of Eq. (4) is monetary compensation payment for shifting device i (a flat payment $C_{A(i)}$ in EUR independent of the periods the device is shifted); the second term is a remuneration given for the modification of the baseline profile of type B devices (where $C_{B(j)}$ is compensation payment in EUR/kWh modification); and the third term is a penalty, C_{DSO} in EUR/kWh, paid for the mismatch between the flexibility procured by the DSO ($F_{DSO(t)}$) and the flexibility provided by the aggregator ($F_{agg(t)}$) in each period t.

Other features and assumptions of the optimization model in track 2 are as follows:

- we consider the perspective of an aggregator connected to home energy management systems with DR capabilities.
- Two types of devices are considered for DR, 1) devices which consumption can be shifted, 2) devices with real-time control capabilities.
- The aggregator has the required infrastructure (e.g., smart metering systems, communication lines, HEMS)to respond to flexibility requests from a distribution system operator who pays monetary compensation for the flexibility.
- the distribution system operator and the aggregator have information (either by a third party or by forecasting tools) of the baseline power consumption (represents normal consumption in case no DR is activated).
- End-users can register and configure the devices for flexibility provision, programming preferences regarding allowed shiftable times, expected remuneration, the priority of the available devices, among others.

More details about the formulation of the problem is available in the publication [7].

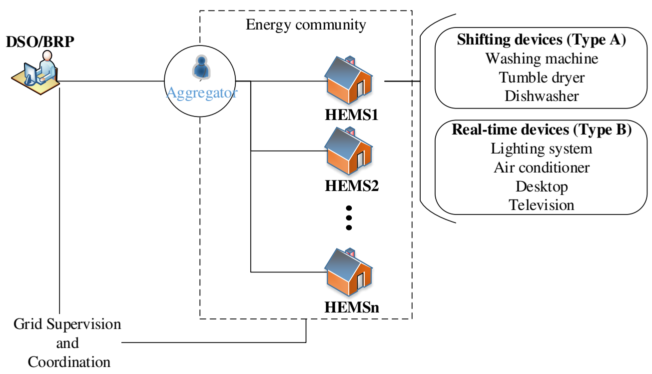

Fig. 2. Track 2 considers a distribution network in which an aggregator is in control of the management of devices with DR capabilities.

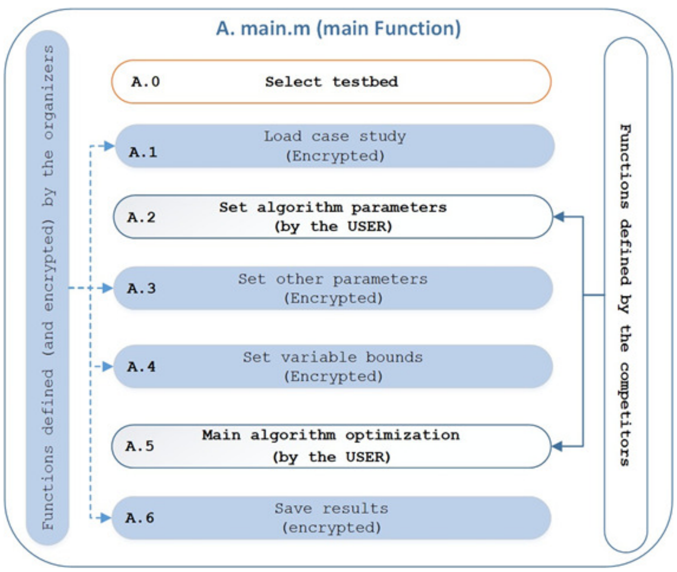

Fig. 3. General framework of the simulation platform - Competition 2021.

III. Simulation Framework and Algorithms

As in previous editions of these competitions, we set a platform where participants can easily implement their algorithms (e.g., metaheuristics such as differential evolution, particle swarm optimization, etc.). We provide an algorithm sample, the HyDE-DF algorithm [10], that can work as a baseline approach. The simulation framework follows the structure of Fig. 3.

We provide a simulation platform implemented in MATLAB© 2018 64-bit. The platform is composed of different scripts that have different functions in the simulation. In Fig. 3, we highlight in blue some encrypted scripts that the organizers use to load the case study (depending on the selected track), set some specific parameters and variable bounds for the simulation, and save automatically the results obtained by participants. The idea is to limit the possibility of changing the case studies by the participants and turning the problems into black-boxes optimization functions.

With these considerations, the participant only needs to implement two scripts: i) one script for setting the parameters required by their algorithm (A.2); ii) a second script for the implementation of their proposed solution method (A.6). A detailed explanation on how to implement these two script functions and how the organizer's scripts work in the platform are provided in [11], Sect. 4.

In the guidelines document [11], we also provide information about the encoding of solutions, assumptions, and the evaluation process. One important parameter to consider is the maximum allowed number of function evaluations on each track, set to 10,000 function evaluations for track1 and 100,000 function evaluations for track2. This is the limit that each participant needs to consider when designing their algorithms, taking into account that a different number of functions evaluations could be done at each iteration depending on the algorithm.

In this competition, we have received the participation of 20 entries, distributed in the different tracks. Table I summarizes the competition entries, which encompasses a nice set of evolutionary algorithms.

IV. Statistical test: Wilcoxon pairwise comparison

The evaluation in this competition has been based solely on the average value of the fitness function over the 20 requested trials:

$$\mathrm{RI}_a^{\mathrm{user}} = \frac{1}{N_{\mathrm{trials}}} \cdot \left[\sum_{i=1}^{N_{\mathrm{trials}}} \mathrm{Fit}_a(\vec{X}_i) \right] \tag{5}$$

where $\mathrm{RI}_a^{\mathrm{user}}$ is the so-called ranking index of participant a, N_{trials} is the number of trials (20 for this competition, and the same number for both tracks and for all participants), and $\mathrm{Fit}_a(\vec{X}_i)$ is a function that receives solution $\vec{X}_i$ and return the fitness value (notice that depending on the track, this function computes the fitness value accordingly). The evaluation of the performance of algorithms is easy and guarantees that the winner approach has achieved a better average fitness on a given track. However, due to the stochastic nature of the solutions provided by these algorithms, it is impossible to claim that a better average fitness guarantees a consistent and statistically better performance of an algorithm with a certain interval of confidence.

An alternative to compare the performance of EAs using statistical tools is the so-called pairwise comparisons. This type of comparison is the simplest kind of statistical test applied within a framework like the one proposed in this work. The test represents a direct comparison between two algorithms that solve a common problem. In this work, we apply the Wilcoxon signed-rank test to compare the performance of the winner approaches. The Wilcoxon test is used to analyze if two samples represent different populations. In other words, the Wilcoxon test is a nonparametric procedure that detects the significant differences (performance or behavior) of two algorithms.

TABLE I
2021 PARTICIPANTS: 20 ALGORITHM SUBMISSIONS FROM A DIVERSIFIED NUMBER OF TEAMS AND COUNTRIES SUBMITTED THEIR RESULTS USING CLASSIC AND HYBRID APPROACHES.

ID	Algorithm	Affiliation	Country	Track
1	Levy Fast Covariance Matrix Adaptation Evolution Strategy (LFC-MAES)	CHARUSAT	India	1 & 2
2	Fast Covariance Matrix Adaptation Evolution Strategy (FC-MAES)			1 & 2
3	Fast Matrix adaptation Evolution Strategy (fastMAES)			1
4	First Coordinate Improvement Evolution Strategy (FCI_ES)			2
5	Cooperative Co-evolution Strategies with Time-dependent Grouping (CCS-TG)	South China University of Technology	China	1 & 2
6	Harris Hawks Optimization + Differential Evolutionary Particle Swarm Optimization + Hybrid-Adaptive Differential Evolution with Decay Function (HHO-DEEPSO-HyDE-DF)	Universidad Nacional de Colombia	Colombia	1 & 2
7	Contribution-Based Cooperative Co-evolution Recursive Differential Grouping (CBCC-RDG3)	St. Petersburg State University	Russia	2
8	Population REgeneration STar-guided Optimization (Presto)	Deakin University	Australia	1 & 2
9	differential evolution with Teaching-Learning-Based Optimization (DE-TLBO)	Sardar Vallabhbhai National Institute of Technology	India	1
10	Genetic Algorithm Simulate Anneling Particle Swarm Optimization (GASAPSO)			1
11	Hill Climbing to Ring Cellular Encode-Decode Univariate Marginal Distribution Algorithm (HC2RCEDUMDA)	Unidad de Transferencia Tecnológica Tepic del Centro de Investigación Científica y de Educación Superior de Ensenada	Mexico, Cuba	1 & 2
12	Artificial Bee Colony (ABC)	Sardar Vallabhbhai National Institute of Technology	India	1 & 2
13	Simulated Annealing Genetic Algorithm Particle Swarm Optimization (SaGaPSO)		India	1 & 2
14	Genetic Algorithm with Particle Swarm Optimization (GA-PSO)		India	1 & 2
15	First Coordinate Improvement Evolution Strategy and Enhance Levy Particle Swarm Optimization (FCI_ES-ELPSO)	CHARUSAT	India	2
16	Gaining Sharing Knowledge - Influence Factor (GSK-IF)	National University of San Luis	Argentina	1 & 2
17	Cellular UMDA with Normal-Gamma distribution (CUMDANGamma)	Camagüey University	Cuba, Mexico	1
18	Cellular UMDA with Normal distribution (CUMDANSimple)	Camagüey University	Cuba, Mexico	2
19	Memory Adaptive Differential Evolution (MJADE)	Ene Operador Regional	El Salvador, USA	1 & 2
20	Success-History based Adaptive Differential Evolution (SHADE)	University of Salamanca	Spain	1 & 2

To define the Wilcoxon test, let d_i be the difference between the performance scores (i.e., the fitness value) of the algorithms over a given problem (i.e., a specific track). Next, these differences are ranked according to their absolute value (there are different methods to deal with ties in the literature [12]; in this work, we used the build-in "tiedrank" function of MATLAB). After that, let R^+ be the sum of ranks in which the first algorithm outperforms the second, and R^- the opposite. Ranks of 0 are evenly split among the sum, and odd numbers are ignored:

$$R^+ = \sum_{d_i>0} \text{rank}(d_i) + \frac{1}{2}\sum_{d_i=0} \text{rank}(d_i) \tag{6}$$

$$R^- = \sum_{d_i<0} \text{rank}(d_i) + \frac{1}{2}\sum_{d_i=0} \text{rank}(d_i) \tag{7}$$

Let T be the minimum value of the rank sums $T = \min(R^+, R^-)$. With these computations, if T is less than or equal to the value of the distribution of Wilcoxon for n degrees of freedom ([13], Table B.12), the null hypothesis H0 of equality of means is rejected (The null hypothesis H0 for this test is: "There is no difference between the median of the solutions of algorithm A and the median of the solutions of algorithm B for the same benchmark problem".). By rejecting the null hypothesis, we imply that a given algorithm outperforms the other one, assuming the associated p-value (0.05 in this work). We have used the in-built function

”signrank” implemented in MATLAB, although other well-known statistical softwares have implemented methods to perform this test.

V. Results and Discussion

In this section, we present the main results of this edition of the competition. We start by verifying the statistical meaning of the RI produced in the competition according to the organizers’ guidelines. We use a confidence level of 95% for these results. The Wilcoxon test as described in Section IV is used for this purpose. The test results depicted in the tables, including the symbols ’+’, ’-’, ’=’, should be interpreted by comparing the top algorithm with the ones in the left. The meaning of ’+’,’-’,’=’ is: significantly better, significantly worse, and not significantly better neither, respectively.

Table II depicts the Wilcoxon test results by comparing all the submitted algorithms in Track 1 with the announced winners of the respective track. It can be seen that CCS-TG is significantly better than the other algorithms that enrolled in the competition. On the other hand, HC2RECEDUMDA is not significantly better than PRESTO, GSK-IF and MJ-DAE, and it is significantly worse than CCS-TG as expected. Regarding PRESTO, it is not significantly better than HC2RECEDUMDA, GSK-IF and MJADE, and it is significantly worse than CCS-TG. Table III depicts the Wilcoxon test results by comparing all the submitted algorithms in Track 2 with the announced winners of the respective track. It can be seen that HC2RECEDUMDA is significantly better than the other algorithms that enrolled in the competition. CCS-TG is better than the other algorithms with one exception; it is significantly worse than HC2RECEDUMDA as expected. Regarding GSK-IF, it is significantly worse than CCS-TG and HC2RECEDUMDA. Tables IV and V show a summary of the results between the winners of each track, respectively. The results clarify that the competition winners in each track are significantly better than the others in their own track. The same cannot be concluded to the second and third place in track 1 since some results come as a surprise. Indeed HC2RECEDUMDA and PRESTO are not significantly better than GSK-IF in track 1. Also, HC2RECEDUMDA and PRESTO compared are not better in statistical terms. In fact, GSK-IF has ranked third in track 2 ahead of PRESTO and the statistical Wilcoxon test shows it is significantly better than it.

The convergence of the top 3 algorithms for track 1 and 2 are depicted in Figs. 4 and 5, respectively. The fitness values have been obtained with a random run of each algorithm and the variability has been confirmed to be robust, i.e., the results are similar in each run. It can be seen that H2RCDEDUMDA presents an initial high convergence rate in both tracks. PRESTO is noticeably slower in track 1. Despite CCS-TG being slower to converge in track 1, it manages to escape from a local optimum where H2RCDEDUMDA gets stuck and obtain a better result with statistical significance, as seen before. In track 2, the H2RCDEDUMDA is clearly the quickest to converge and the best in the trials. GSK-IF converges very slowly. The convergence properties suggest that if the competition allowed more fitness function evaluations, GSK-IF could become the ”winner”, at least in track 2.

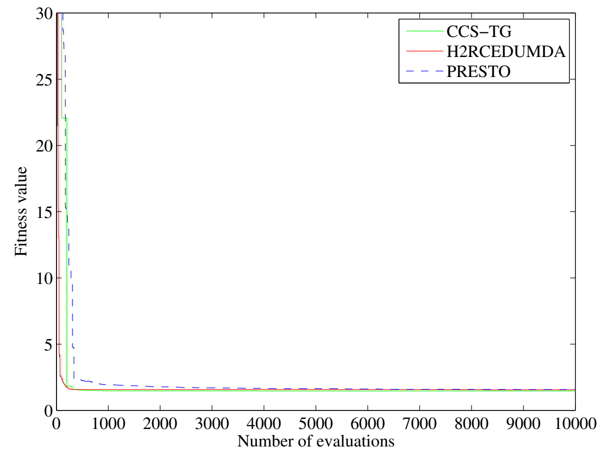

Fig. 4. Convergence properties of the top 3 ranked algorithms in Track 1.

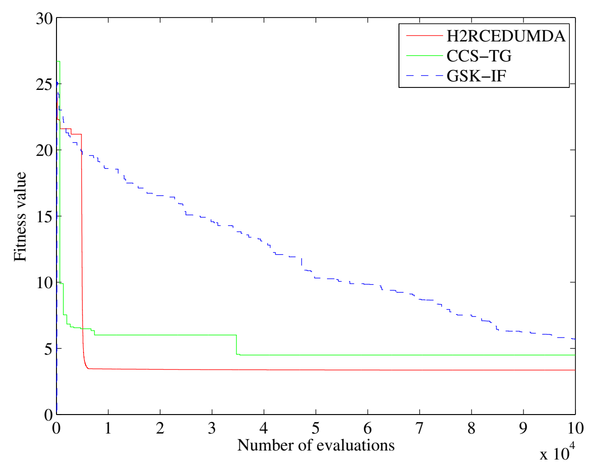

Fig. 5. Convergence properties of the top 3 ranked algorithms in Track 2.

VI. Conclusions

This edition of this competition in the energy domain provides a platform to test and compare new EAs to solve complex problems in the field. The winning algorithms of this edition, CCS-TG in track 1 and H2RCDEDUMDA in track 2, were significantly better than the other participants. These algorithms were clear winners in their own track since their results are statistically significant when compared with the other entries. Therefore the ranking index worked well for those cases. Regarding second place and third place, in track 1, our analysis raised some doubts about the statistical meaning of their differences. Their differences are not very meaningful in 20 trials as tested. This issue could perhaps be investigated by running more than 20 trials as it was initially established in our competition rules. We would suggest at

TABLE II
WILCOXON TEST ALL PARTICIPANTS TRACK 1

	CCS-TG (1st)	**HC2RCEDUMDA (2nd)**	**PRESTO (3rd)**	**RI (position)**
LFC-MAES	'+'	'+'	'+'	1.933 (14th)
FC-MAES	'+'	'+'	'+'	1.895 (13th)
fastMAES	'+'	'+'	'+'	1.637 (8th)
CCS-TG	**'='**	**'-'**	**'-'**	**1.478 (1st)**
HHO-DEEPSO-HyDE-DF	'+'	'+'	'+'	1.636 (7th)
GASAPSO	'+'	'+'	'+'	1.659 (9th)
HC2RCEDUMDA	**'+'**	**'='**	**'='**	**1.544 (2nd)**
ABC	'+'	'+'	'+'	1.801 (12th)
SaGaPSO	'+'	'+'	'+'	1.659 (9th)
GAPSO	'+'	'+'	'+'	1.772 (11th)
DE-TLBO	'+'	'+'	'+'	2.035 (15th)
PRESTO	**'+'**	**'='**	**'='**	**1.548 (3rd)**
GSK-IF	'+'	'='	'='	1.552 (5th)
CUMDANGamma	'+'	'+'	'+'	1.567 (6th)
MJADE	'+'	'='	'='	1.549 (4th)
SHADE	'+'	'+'	'+'	1.687 (10th)

TABLE III
WILCOXON TEST ALL PARTICIPANTS TRACK 2

	HC2RCEDUMDA (1st)	**CCS-TG (2nd)**	**GSK-IF (3rd)**	**RI (position)**
LFC-MAES	'+'	'+'	'+'	10.896 (16th)
FC-MAES	'+'	'+'	'+'	10.776 (15th)
FCI_ES	'+'	'+'	'+'	7.660 (10th)
CCS-TG	**'+'**	**'='**	**'-'**	**4.349 (2nd)**
HHO-DEEPSO-HyDE-DF	'+'	'+'	'+'	6.919 (5th)
CBCC-RDG3	'+'	'+'	'+'	8.316 (12th)
HC2RCEDUMDA	**'='**	**'-'**	**'-'**	**3.493 (1st)**
ABC	'+'	'+'	'+'	7.967 (11th)
SaGaPSO	'+'	'+'	'+'	7.341 (8th)
GAPSO	'+'	'+'	'+'	7.643 (9th)
FCI_ES-ELPSO	'+'	'+'	'+'	7.140 (6th)
PRESTO	'+'	'+'	'+'	5.789 (4th)
GSK-IF	**'+'**	**'+'**	**'='**	**4.795 (3rd)**
CUMDANSimple	'+'	'+'	'+'	8.484 (13th)
MJADE	'+'	'+'	'+'	7.146 (7th)
SHADE	'+'	'+'	'+'	10.429 (14th)

TABLE IV
WILCOXON TEST BETWEEN THE WINNERS TRACK 1

	CCS-TG (1st)	**HC2RCEDUMDA (2nd)**	**PRESTO (3rd)**
CCS-TG (1st)	'='	'-'	'-'
HC2RCEDUMDA (2nd)	'+'	'='	'='
PRESTO (3rd)	'+'	'='	'='

TABLE V
WILCOXON TEST BETWEEN THE WINNERS TRACK 2

	HC2RCEDUMDA (1st)	**CCS-TG (2nd)**	**GSK-IF (3rd)**
HC2RCEDUMDA (1st)	'='	'-'	'-'
CCS-TG (2nd)	'+'	'='	'-'
GSK-IF (3rd)	'+'	'+'	'='

least 100 trials to verify this issue in future work. In fact, another surprise is that GSK-IF and MJADE are equivalent to H2RCDEDUMDA (2nd place) and PRESTO (3rd place). As a result of this, if the metric in the competition would have been the Wilcoxon test, there would be three algorithms in the second place. Another highlight of our analysis is that GSK-IF convergence seems to be slower. Hence, if more evaluations were given, GSK-IF could perhaps provide better solutions than the winning algorithms. This is an aspect that can be investigated in future work. Future editions of this competition should look to include some statistical meaning metric instead of the useful but simple ranking index as in the past editions.

REFERENCES

[1] P. Siano, "Demand response and smart grids—a survey," *Renewable and sustainable energy reviews*, vol. 30, pp. 461–478, 2014.

[2] J. Beck, R. Kempener, B. Cohen, and J. Petrie, "A complex systems approach to planning, optimization and decision making for energy networks," *Energy policy*, vol. 36, no. 8, pp. 2795–2805, 2008.

[3] J. Soares, T. Pinto, F. Lezama, and H. Morais, "Survey on complex optimization and simulation for the new power systems paradigm," *Complexity*, vol. 2018, 2018.

[4] F. Lezama, J. Soares, Z. Vale, J. Rueda, S. Rivera, and I. Elrich, "2017 ieee competition on modern heuristic optimizers for smart grid operation: Testbeds and results," *Swarm and evolutionary computation*, vol. 44, pp. 420–427, 2019.

[5] A. Kumar, G. Wu, M. Z. Ali, R. Mallipeddi, P. N. Suganthan, and S. Das, "A test-suite of non-convex constrained optimization problems from the real-world and some baseline results," *Swarm and Evolutionary Computation*, vol. 56, p. 100693, 2020.

[6] F. Lezama, J. Soares, R. Faia, Z. Vale, O. Kilkki, S. Repo, and J. Segerstam, "Bidding in local electricity markets with cascading wholesale market integration," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107045, 2021.

[7] F. Lezama, J. Soares, B. Canizes, and Z. Vale, "Flexibility management model of home appliances to support dso requests in smart grids," *Sustainable Cities and Society*, vol. 55, p. 102048, 2020.

[8] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3 – 18, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210650211000034

[9] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997.

[10] F. Lezama, J. Soares, R. Faia, T. Pinto, and Z. Vale, "A new hybrid-adaptive differential evolution for a smart grid application under uncertainty," in *2018 IEEE Congress on Evolutionary Computation (CEC)*, July 2018, pp. 1–8.

[11] F. Lezama, J. Soares, B. Canizes, Z. Vale, and R. Romero, "Guidelines for pes-gm/cec/gecco 2021 competition evolutionary computation in the energy domain: Smart grid applications," 2021.

[12] J. D. Gibbons and S. Chakraborti, *Nonparametric statistical inference: revised and expanded*. CRC press, 2014.

[13] J. H. Zar, *Biostatistical analysis*. Pearson Education India, 1999.