

A hybrid swarm-based algorithm for single-objective optimization problems involving high-cost analyses

*Original*

A hybrid swarm-based algorithm for single-objective optimization problems involving high-cost analyses / Ampellio, Enrico; Vassio, Luca. - In: SWARM INTELLIGENCE. - ISSN 1935-3812. - 10:2(2016), pp. 99-121. [10.1007/s11721-016-0121-6]

*Availability:*

This version is available at: 11583/2641060 since: 2017-04-27T18:59:02Z

*Publisher:*

Springer New York LLC

*Published*

DOI:10.1007/s11721-016-0121-6

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

Springer postprint/Author's Accepted Manuscript

This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <http://dx.doi.org/10.1007/s11721-016-0121-6>

(Article begins on next page)

# A hybrid swarm-based algorithm for single-objective optimization problems involving high-cost analyses

Enrico Ampellio · Luca Vassio

Received: May 26, 2015 / Accepted: date

**Abstract** In many technical fields, single-objective optimization procedures in continuous domains involve expensive numerical simulations. In this context, an improvement of the *Artificial Bee Colony* (ABC) algorithm, called the *Artificial super-Bee enhanced Colony* (AsBeC), is presented. AsBeC is designed to provide fast convergence speed, high solution accuracy and robust performance over a wide range of problems. It implements enhancements of the ABC structure and hybridizations with interpolation strategies. The latter are inspired by the quadratic trust region approach for local investigation and by an efficient global optimizer for separable problems. Each modification and their combined effects are studied with appropriate metrics on a numerical benchmark, which is also used for comparing AsBeC with some effective ABC variants and other derivative-free algorithms. In addition, the presented algorithm is validated on two recent benchmarks adopted for competitions in international conferences. Results shows remarkable competitiveness and robustness for AsBeC.

**Keywords** modified Artificial Bee Colony · engineering optimization · interpolation strategies · algorithm comparison

## 1 Introduction

Numerical optimization in applied sciences usually concerns simulation-based problems in which computational analyses are expensive in terms of time and resources (S. Rao, 2009; K. Deb, 2012). In engineering, complex non-linear models are commonly used (Lagaros and Papadrakakis, 2015), e.g., computational fluid dynamics and finite element methods. This field demands fast algorithms for solving optimization problems using as few function evaluations (FEs) as possible. The present work focuses on single-objective problems with low, but not minimal dimensionality.

In the literature, there are several architectures and algorithms suitable for global optimization in tough numerical problems (Floudas and Pardalos, 2009). Among the most studied are surrogates (also called meta-models), heuristic and statistical methods, trust region approaches and nature-based algorithms.

Surrogate models (Kozziel et al., 2013) are effectively exploited in a large class of real-world applications. They are often the first choice and sometimes the only practical one when very expensive analyses are involved. Response surfaces (Box and Draper, 2007), artificial neural networks (Iliadis and Jayne, 2015), radial

---

E. Ampellio (✉)

Dipartimento di Ingegneria Meccanica e Aerospaziale, Politecnico di Torino, 10129 Torino (TO), Italy

E-mail: enrico.ampellio@polito.it

L. Vassio

Dipartimento di Elettronica e Telecomunicazioni, Politecnico di Torino, 10129 Torino (TO), Italy

E-mail: luca.vassio@polito.it

basis functions (Buhmann, 2003) and Kriging (Simpson et al., 2001) are surrogates widely adopted to support high computing load optimizations. One of the first optimizers fully based on meta-models is the Efficient Global Optimization (Jones et al., 1998). This algorithm and its successors are expressively targeted to problems where function evaluations are strictly limited by time or cost. Heuristic and statistical techniques such as Sequential Parameter Optimization (SPO, Bartz-Beielstein et al., 2005) are also related to meta-models and intensively used in applicative fields. They are focused on selecting and tuning both the optimization methodology and the search methods, but they imply that efficient algorithms and/or surrogates for a specific problem are a priori available. Nonetheless, ensuring the reliability of surrogates is still a sensitive issue and the training of meta-models could not be cost-effective when considering highly non-linear, noisy, multi-modal and discontinuous problems. Moreover, this approach becomes computationally impracticable when increasing the problem dimensionality. As a result, it is attractive to develop direct search algorithms designed for costly optimizations, especially to cope with higher dimensions (e.g., 10 dimensions).

Among direct search algorithms, derivative-free trust region optimizers (Conn et al., 2000), such as the quadratic UOBYQA (Powell, 2000) and related approaches (e.g., CONDOR by Berghen and Bersini, 2005; BOBYQA by Powell, 2009), are effective for heavy computing loads and noisy objective functions. The hybridization with quadratic local meta-models is also used to tackle costly problems, for instance in evolutionary strategies (Imm-ES by Kern et al., 2006), in which surrogates are periodically called to improve performance and save expensive function evaluations (FEs).

Meta-heuristic direct-search algorithms play a predominant role when dealing with single-objective optimizations that comprise many continuous variables and noisy multi-modal target functions, and where analytical properties such as differentiability are not known a priori (Talbi, 2009). Several nature-based techniques have been developed such as genetic algorithms (Goldberg, 1989), particle swarm optimization (Kennedy and Eberhart, 1995), ant colony optimization (Dorigo and Stützle, 2004), differential evolution (DE by Price et al., 2005) and evolutionary strategies (Beyer and Schwefel, 2002).

One of the more recent and promising nature-inspired meta-heuristic methods is the Artificial Bee Colony algorithm (ABC) by Karaboga and Akay (2007). ABC has been compared with other nature-inspired algorithms and it showed good results for a great variety of optimization problems (Karaboga and Akay, 2009) as well as for real-world applications (Akay and Karaboga, 2010). Besides, a lot of test case validations and advancements have been done in recent years (Bolaji et al., 2013). The present paper further improves the accuracy and convergence speed of ABC algorithms to make them suitable for continuous optimization problems with expensive analyses. This means obtaining better solutions using the same maximum number of function evaluations. The main idea behind the proposed algorithm is to hybridize ABC with the basic principles of derivative-free, quadratic trust region optimizers for high-cost objective functions. Moreover, another hybridization with the fast and robust S.T.E.P method (Swarzberg et al. 1994) is introduced to efficiently cope with separable functions. In the end, AsBeC aims at showing robust performance over a wide range of problems by successfully merging the main advantages of swarm-based and interpolative methods.

To introduce the main peculiarities of ABC, a brief review of the original algorithm and its popular modifications is presented in Section 2. The proposed enhancements and hybridizations leading to the novel Artificial super-Bee enhanced Colony algorithm (AsBeC) are then discussed in Section 3. The analytical benchmarks and new metrics used for comparisons are respectively introduced in Section 4 and Section 5. The performance of the new algorithm is accurately evaluated and compared to ABC through extensive simulations in Section 6. In Section 7, some ABC variants and other state-of-the-art algorithms for direct search and appropriate for costly optimizations are considered for comparisons. In Section 8, the algorithm is validated using two other famous benchmarks for expensive optimization. Conclusions and insights about future work on the topic are presented in Section 9. An electronic appendix gathers all comparative results in expanded form.

## 2 ABC and its modifications in the literature

The Artificial Bee Colony (ABC) algorithm by Karaboga (2007) is an example of swarm intelligence as it models the collective behaviour of decentralized and self-organized systems consisting of simple agents. The algorithm reproduces the behaviour of a honey bee colony searching the best nectar source in a target area.

Some bees, called the employees, are each assigned to a nectar source and search the space near it. Then, they come back to the hive and communicate to other bees, called the onlookers, the position of the best nectar sources found through a dance (Frish, 1967), so that onlookers can help the employed bees in searching the most promising regions. Nectar sources that reveal themselves non-productive are abandoned in favour of new positions randomly chosen in the search domain, which are investigated by scout bees.

In the optimization context, nectar sources represent the seed solutions and the nectar amount corresponds to the objective function value to optimize; "non-productive" sources are those not improving for some time, controlled by the limit parameter. An entire optimization sequence composed by the employee group, performing the exploration, the onlooker group, performing the exploitation, and the scout bee, maintaining the population diversity, is called a *cycle*. For an ABC pseudo-code refer to Section 2.1 in Karaboga and Akay (2007).

The basic search movement for a seed solution  $x^j$  is called mutation equation and it modifies only one variable at a time,  $i$ , chosen uniformly at random. Another seed solution  $x^k \neq x^j$  is also chosen uniformly at random and the new candidate solution  $x^{j,rnd}$  in variable  $i$  is:

$$x_i^{j,rnd} = x_i^j + rnd[-1, 1] \cdot (x_i^j - x_i^k) \quad (1)$$

where  $rnd[-1, 1]$  is a uniformly distributed random number in the continuous range  $[-1, 1]$ .

The ABC algorithm tries to balance exploration and exploitation behaviours, offering worthy global and local search skills at once. As a result, when compared with other methods, ABC demonstrates high-performance and suitability for a broad spectrum of optimization problems. To enter in details consult Karaboga (2007) and for extensive simulation comparisons refer to Karaboga and Akay (2009). According to many authors, ABC algorithm is simple to implement, easy to be effectively parallelized (e.g., Subotic et al., 2011) and hybridized, driven by few control parameters, flexible and robust (e.g., Karaboga et al. 2014). On the other hand, its semi-random movement is not taking into account the local shape of the function. Moreover, the algorithm does not exploit the history of the tested solutions.

Since the original paper by Karaboga (2007), many researches were developed to overcome the deficiencies of the original ABC. In particular, many scientists focused on the common goal to advance the local search mechanism without worsening the global one. In this regard, specific modifications of various nature have proven to be convenient (Bolaji et al., 2013). Among them, the popular ABC variants by Gao and Liu (e.g., IABC, 2011 and BABC, 2012) and by Kong et al. (e.g., 2013) are focused on initialization enhancements and DE/PSO inspired mutation equations. Nevertheless, many preliminary tests performed by the authors of the present paper showed that these modifications are not of major impact when dealing with a limited number of FEs and small colonies. Their performance in this framework is comparable with the simpler and well-known GABC (Zhu and Kwong, 2010), which drives the classic bee movement towards the global best solution.

Some other variants give more importance to the exploitation attitude by altering the fitness function or the assignment methodology for onlookers (e.g., Subotic, 2011). Among them, JA-ABC5 (Sulaiman et al., 2015) introduces a multi-stage re-organization of the bees which seems particularly beneficial.

Many modifications have been proposed coupling ABC either with classical optimization methods for local search such as simplex (HSABC by Kang et al., 2009), the Rosenbrock method (RABC by Kang et al., 2011) and pattern search (HJABC by Kang et al., 2013), or with other robust and efficient well-known techniques for global search, such as genetic algorithms (GA-ABC by Zhao et al., 2010) or evolutionary strategies (BE-ABC by Li and Li, 2012). Among these methods, the RABC based on the Rosenbrock directional search (Rosenbrock, 1960) shows fast convergence to the global optimum.

Despite the great amount of available literature on ABC variants, to the best of our knowledge no paper underlines a performance gain even with very few function evaluations. As well, no hybridization with methods similar to trust region techniques has been analysed so far, even though they are recognized to be among the fastest techniques, especially when dimensionality is small (Rios and Sahinidis, 2013). These reasons justify the introduction and analysis of ABC modifications specifically targeted to costly optimization and real-world oriented problems.

### 3 Improving techniques for the AsBeC algorithm

Among all the proposed modifications of the ABC algorithm in the literature, some techniques are on average more effective or more robust (Liao et al., 2013 and Aydin, 2015) for specific problems. In the present work, the authors discover that modifying the onlooker assignment stage, exploiting the data history according to the theory of collective memory in social animals (Couzin et al., 2002) and applying some hybridizations like local interpolation are especially beneficial when the number of FEs is very limited. The presented techniques have the purpose of achieving the fastest solution improvement, without clustering the swarm and leading to premature convergence at the same time.

The sections below describe each technique in details. A minimization problem is considered without loss of generality. The final AsBeC pseudo-code is outlined and commented in Section 3.6.

#### 3.1 Strictly biased onlooker assignment

In ABC, onlookers are assigned to seed solutions by a stochastic rule. A certain probability is associated to each of them: it is proportional to the solution quality  $f$ , through a fitness function  $fit$ . Probability and fitness definitions are taken from Akay and Karaboga (2012). However, if the value of  $f$  for all the seed solutions is similar, that fitness function returns almost identical values and the relative probabilities are alike. Hence, the standard fitness formulation is not able to distinguish the solution quality in relative terms. In order to always strengthen exploitation, it is possible to unbalance onlooker assignment by normalizing the fitness  $fit$  to a new re-scaled function  $fit^{res}$ :

$$fit^{res}(x^k) = \frac{fit(x^k) - \min_{s=1, \dots, SN} fit(x^s)}{\max_{s=1, \dots, SN} fit(x^s) - \min_{s=1, \dots, SN} fit(x^s)} \quad \forall k \in \{1, \dots, SN\}$$

where  $x^k$  is a seed solution and  $SN$  is the number of seed solutions. The above formulation for  $fit^{res}(x^k)$  is generalized and still valid even when changing the definition of  $fit$ . Different from the rank-based fitness transformation of Kang et al. (2011 and 2013), it takes into account also the quality difference among seed solutions. In fact,  $fit^{res}$  scores the seed solutions from 0 to 1 according to their distance from the best.

In the original ABC formulation, there is a possibility to assign few or no bees to the best seed solutions. The previous scenario is unlikely when dealing with a large-size swarm; however, for small-size colonies a different approach should be preferred in order to save function evaluations. A deterministic rule for onlookers assignment is proposed in place of the original stochastic one, in order to associate a prearranged number of bees  $n^k$  to seed solution  $k$  proportionally to its  $fit^{res}$ :

$$n^k = \left\lfloor \frac{ON \cdot fit^{res}(x^k)}{\sum_{s=1}^{SN} fit^{res}(x^s)} \right\rfloor \quad \forall k \in \{1, \dots, SN\}$$

where  $ON$  is the onlooker number and the operator  $\lfloor \cdot \rfloor$  returns the largest integer smaller than the argument. In case the number of onlookers assigned,  $\sum_{k=1}^{SN} n^k$ , is lower than the total number of onlookers  $ON$ , the difference  $ON - \sum_{k=1}^{SN} n^k$  is assigned to the best seed solution. In practice, no bees are ever sent to the worst seed solution and at least one is always sent to the best one. The application of this deterministic rule is similar to the idea of exploiting only few best solutions of the JA-ABC5 algorithm and it always encourages the repositioning near the global best, as in GABC; however, it is self-adapted during the optimization process.

Notice that this strictly biased onlooker (BO) assignment, through fitness re-scaling and deterministic rule, does not appreciably affect the global search skills of the swarm. In fact, exploration is essentially up to the employee movement, which remains unaltered. This technique intentionally disregards the worst areas, similarly to the principle driving other swarm-based algorithms such as the FIREFLY algorithm (Yang, 2009).

### 3.2 Postponed onlooker dance

This modification enables the onlooker bees to seek nectar near the best regions for a longer time. The entire onlooker group performs more than one iteration sequentially, without altering the updating order of seed solutions inside each iteration. As in the original ABC, a seed solution is immediately updated whenever a bee finds an improvement for it. In the postponed onlooker dance (PD) iterations, seed solutions are never reassigned, but their updated versions are the starting configurations for the new movements. In this way, each onlooker bee has multiple chances to improve the seed solution to which it was assigned and to interact with the other proposed techniques. Anyway, the number of PD iterations should be maintained small in order not to affect exploration and decelerate the convergence speed to the global optimum. In this paper, the number of reiterations on the onlooker group is set to three.

This modification acts like enlarging the onlooker group, but repeating their movements instead of increasing their number. Hence, the best seed solutions have a higher probability to improve and to be selected for mutation than the others, as onlookers are not distributed on a larger basis.

This postponed onlooker dance among bee groups is inspired by nature. In fact, bees normally come back to the hive to communicate their seed solutions only after some time, when there is more probability to have collected new pollen. This technique is inspired by the work of Subotic (MO-ABC, 2011), except that the PD proposed is a deterministic procedure.

### 3.3 Local interpolation

In order to quickly improve the best solutions in their local neighbourhood, the concepts of opposition principle and parabolic interpolation are introduced.

The opposition-based learning for the ABC algorithm has already been suggested in the OABC by El-Abd (2011). In the present paper, the following version is proposed: whenever a pseudo-random movement for the seed solution  $k$  to a new position  $x^{k,rand}$  does not produce any improvement, it is possible to move in the same direction with the same step, but in the opposite sense. If this new position  $x^{k,opp}$  is better than the previous, the correspondent seed solution  $k$  is updated. As a result, the opposite principle is a linear local estimator when the step  $\|x^k - x^{k,rand}\|$  is small enough and the objective function is continuously differentiable in the neighborhood of  $x^k$ . The operator  $\|\cdot\|$  expresses the Euclidean norm.

On the other hand, the parabolic interpolation estimates the local and directional curvature of the objective function, acting as a second order optimization method with partial Hessian computation. This technique follows the opposition principle: if the opposite step is not successful, the multi-dimensional parabola passing through three already known points, i.e., (i) the point  $x^k$ , (ii) the point generated by the first random movement  $x^{k,rand}$  and (iii) the opposite point  $x^{k,opp}$ , is calculated and its minimum is tested. The opposition principle and second order interpolation are performed in excluding sequence: if the first random movement is improving, none of the two following steps is carried out; if the opposite movement is improving, the parabola will not be estimated; if the two previous steps lead to solutions worse than the seed solution, then the second order interpolation is computed.

Local interpolation (LI) is used in ABC when at least three onlooker bees are assigned to the same seed solution. When the PD technique is activated, the onlooker group has more chances to exploit local interpolation. During the onlooker phase, the previous bee movements relative to the same seed solution are recorded and shared among bees, also considering different PD iterations. When performing local interpolation, it is possible to use candidate solutions generated by different onlooker bees that are assigned to the same seed solution. Using three iterations for the PD technique, a seed solution with just one onlooker assigned could still perform a complete sequence of random, opposite and parabolic interpolation movements.

The raw approximation offered by local interpolation is balanced by the low number of function evaluations required, i.e., respectively one and two additional FEs for the opposite movement and the parabolic interpolation, regardless of the problem dimensionality.

### 3.4 Quadratic prophet

A hybridization with trust region concepts is introduced in order to improve the convergence speed, especially in the early phases. The quadratic prophet (QP) hybridizes ABC with a fast and robust local search method, which is the same philosophy of other hybrids such as RABC (Kang et al., 2011). The key idea is to exploit the function evaluations already performed to create multiple quadratic polynomial interpolators around the best regions. Hence, these quadratic models are built in the neighborhood of the seed solutions by using their closest samplings in the so-called *hive memory*. This memory consists in all the solutions generated by the AsBeC algorithm so far. The global minima of the quadratic models, if they exist, are then tested. The Imm-CMAES algorithm by Kern et al. (2006) is similar, but it uses the quadratic models as response surfaces to be called in place of direct function evaluations. Instead, QP does not require additional FEs to populate the samplings for the interpolations; its final cost is at the most equal to the number of seed solutions if all quadratic models succeeded. Whenever the QP finds a better solution, the related seed solution is immediately updated.

Although this method does not regulate directly the trust region radius, it adapts itself in accordance to the hive memory. In fact, if the samplings around a given seed solution are dense, the algorithm is exploiting the area, and thus the trust region radius is small in order to refine the local solution. On the contrary, when the samplings are sparse, the trust region radius is large to support exploration. This self-adaptation resembles the mechanism used in the CMA-ES algorithm (Hansen, 2006). Other famous trust region methods, such as the classic Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) or the recent BOBYQA (Powell, 2009) algorithms follow the same basic principle. The main difference is that the QP technique uses the data that are already available within the AsBeC algorithm, instead of adopting a costly and complex population strategy to train and use quadratic approximations. On the other hand, the quality of the QP model depends on the quality of the samplings, hence on the ability of AsBeC to properly investigate the search space surrounding the minima. Therefore, the Quadratic Prophet technique can hybridize any optimization algorithm able to provide well-distributed samplings.

The implementation of QP is based on quadratic polynomial surfaces, whose complete form,  $Q_C$ , and reduced form,  $Q_R$ , for a generic problem in  $D$  dimensions are expressed as:

$$Q_C(x_1, x_2, \dots, x_D) = \sum_{i=1}^D \sum_{j=1}^D a_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^D b_i \cdot x_i + R \quad Q_R(x_1, x_2, \dots, x_D) = \sum_{i=1}^D a_{ii} \cdot x_i^2 + \sum_{i=1}^D b_i \cdot x_i + R$$

where  $a_{ij}$ ,  $b_i$  and  $R$  are the  $C$  coefficients of the model, while  $x_i$  are the independent variables of the problem.  $Q_C$  needs  $C = 0.5 \cdot (D + 2) \cdot (D + 1)$  samplings to be exactly defined, since  $a_{ij} = a_{ji}$ . The reduced quadratic surface  $Q_R$ , without any mixed term, only needs  $2 \cdot D + 1$  samplings. The reduced model is not able to capture rotated functions, but it is enough to cope with bowl shape regions. In order to activate this technique as soon as possible, reduced models  $Q_R$  are built starting when the algorithm provides at least  $2 \cdot D + 1$  solutions, ending up with complete models  $Q_C$  when more than  $C$  samplings are available.

The model coefficients are obtained in the simplest possible way, that is by solving exactly the linear system of equations for the  $C$  samplings nearest to a given seed solution. If the Hessian matrix of the model is diagonal positive, then the point in which the Jacobian determinant is zero is tested. This point could represent either a local minimum, if the Hessian is positive definite, or a saddle point dominated by squared terms. Actually, the latter can be wrongly recognized as a saddle instead of a minimum due to inaccuracies. Nonetheless, it represents a key point to better describe the objective function contour in that region. Anyway, eventual redundant QP predictions are not tested twice, in order to save precious FEs.

Notice that the model could misinterpret the local behaviour of the objective function if the sampling is ill-conditioned or if  $f$  is not quadratic or noisy in that area, as common in real-world problems. Then, if the exact solution does not lead to a point to test, the system is solved by means of an iterative method with large tolerance on the same  $C$  samplings. This alternative is aimed at obtaining quickly a simple-to-handle coarse shape interpreter, passing nearby the sample points and not exactly through them. This inaccurate model could still be able to capture the objective function profile and drive the search towards the most promising regions.

The tolerance for the iterative resolution of quadratic systems is set to  $10^{-1}$  by using the QMR method (Roland and Nachtigal, 1991).

In order to increase the likelihood of a well-conditioned sampling, the points should be heterogeneously different in their coordinates. There are two main features that affect the sampling diversity: the standard bee movement, changing only one variable at a time, and the LI technique, producing sequentially aligned points. Therefore, when the QP is activated, LI is deactivated and the mutation equation is slightly modified. Following the idea introduced by Akay and Karaboga (2012) for the MABC, more than one variable is chosen when moving a bee, according to the standard ABC equation. The number of variables to be changed concurrently is set as half of the total number of dimensions, i.e.,  $D/2$ . Clearly, these two modifications overweight exploration and hence have to be restricted to the early phases of the optimization. In this work, they are adopted while the total number of solutions tested is lower than  $2 \cdot C = (D + 1) \cdot (D + 2)$ . In this way, a large pool of samplings more likely to be well-conditioned is available for complete quadratic models.

This QP technique is a finer exploitation of the seed solutions and, hence, it is implemented as an additional exploitative stage at every cycle, after the onlooker phase. It is also used once during the initialization. Indeed, the number of random initial evaluations is set to be (at least) equal to  $2 \cdot D + 1$ , and a fully reduced  $Q_R$  is already feasible.

It must be clear that the QP is designed to deal with problems in which the objective function evaluation represents the bottleneck of computational times. In fact, the computational cost of this technique is dominated by the resolution of the linear system, that in its general form requires  $O(C^3)$  operations at each cycle, for each food source. Given that  $C$  increases along with  $D$ , the technique becomes numerically heavy when addressing high dimensionality problems.

### 3.5 Systematic global optimization

The S.T.E.P algorithm by Swarzberg et al. (1994) shows high convergence rate for highly multi-modal and complex functions of one variable. It is derivative-free, not population-based, without parameters to tune and it does not assume any property for the function to optimize. It relies on the history of the points tested during its functioning. The domain boundaries and a random point between them, called the *context*, are evaluated. These three points define two *partitions*. For each partition is computed the curvature a parabola should have to enclose the best so far solution inside it. This curvature is called the partition *difficulty*. The S.T.E.P method iterates by sampling the center of the partition with smallest difficulty, i.e. the one identified with the greatest chance of improving the best solution found so far, until a given tolerance is reached. The S.T.E.P. authors showed that around 50 FEs are sufficient to solve difficult univariate problems, such as Michalewicz's second function. The simplest multivariate extension to the S.T.E.P. algorithm solves one dimension at a time. The recent version by Baudiš and Pošík (2015) interleaves the steps of the univariate solvers such that all dimensions are optimized concurrently, in a similar way to Rosenbrock algorithm (1960). Each time, the dimension to investigate is chosen following a round-robin scheduling.

Given the fast convergence and robust performance of the S.T.E.P. algorithm, it can be hybridized with ABC, which is more oriented to exploration. However, in a multivariate environment S.T.E.P. based solvers are very effective only on additively separable or quasi-separable problems, in which the correlations among variables are weak or only few variables are correlated. As a consequence, this technique is activated if a test for quasi-separability is successfully passed. The proposed test investigates the separability in the cheapest possible way, giving just a necessary but not sufficient condition. For an additively separable function, each first-order partial derivative  $\partial f / \partial x_i$  is independent from other variables  $x_j \neq x_i$ , giving  $\partial^2 f / \partial x_i \partial x_j = 0$ . For each design variable  $x_i$ , another variable  $x_j$  with  $i \neq j$  is randomly chosen to assess the correlation between them. Then, the partial derivative  $\partial f / \partial x_i$  is estimated for two different values of  $x_j$ , while all other coordinates are kept constant. If the two estimations vary below a given tolerance, here chosen as  $10^{-3}$ , the two variables could be quasi-separable. The same check is repeated for  $\partial f / \partial x_j$  for two different values of  $x_i$ . In order to estimate the aforementioned derivatives through variations, four FEs are needed. These four solutions are chosen around the active seed solutions. If the non-separability condition on the derivatives holds for  $D$  couples of variables, the test is passed. The procedure is interrupted whenever a variable is found to be correlated. This test needs at



most  $2 \cdot D + 1$  function evaluations, since a complete set of partial derivatives in the same point can be estimated only once and reused for all the couples.

If this test is successfully passed, the interleaved solver by Baudiš and Pošík (2015) is run as a one-shot technique after the initialization phase, using the best so-far solution as context and investigating all the variables, in deterministic sequence. For this reason, the proposed hybridization is called systematic global optimization (SGO). At the end of the SGO phase, the best seed solution is moved to the best solution found. The maximum number of evaluations per variable is set to 50, according to the results by Swarzberg et al. (1994). Then, this technique along with its test uses up to  $52 \cdot D + 1$  function evaluations.

### 3.6 The AsBeC algorithm

The combination of all the previous techniques, integrated into ABC, gives the AsBeC algorithm. Its pseudocode is presented in Algorithm 1, indicating the sections in which the new techniques added to ABC are described. BO and PD are enhancements of ABC regarding bees re-organization, while LI, QP and SGO are hybridizations to estimate the local and global behaviour of the objective function, without changing the mutation equation itself. Thanks to these hybridizations, the bees assume new extra abilities and are called super-bees. This explains the name of the new algorithm: the Artificial super-Bee enhanced Colony (AsBeC).

Actually, the authors experimented also other modifications, e.g. in the mutation equation, and other hybridizations with derivative-free directional search techniques, like the discrete gradient method as in Bagirov et al. (2008). In the end, only the variants recognized to have an effective role for the specific goals of this work were implemented within AsBeC.

## 4 Benchmarks

Three different benchmarks for unconstrained problems in the continuous domain are adopted in this paper. The first, Set A, is used to test AsBeC techniques, to provide standard settings for the new algorithm and to compare AsBeC with ABC variants and other notable optimizers. The other two, Set B and C, are popular benchmarks used for competitions in international conferences, herein presented for comparing AsBeC with the most recent methods presented at the CEC 2015 and GECCO 2015 conferences.

### 4.1 Set A

Set A contains unimodal, multimodal, separable and non-separable functions selected among classic analytical problems. Their definitions, search domains and main characteristics are described in Table 1. They are all non-negative and have the global minimum value exactly equal to zero. Functions without an exact analytical characterization of the optimum, e.g., the Styblinski-Tang, are not considered in order to avoid bias on results. Furthermore, all the functions chosen are scalable to any dimension of the variable space. For a rather complete summary of classical analytical functions used for global optimization refer to Jamil and Yang (2013). Almost all functions in Table 1 are very popular among ABC modifications and usually adopted for this kind of assessment (e.g., Gao and Liu, 2011).

Set A is heterogeneous and relatively simple, since it does not contain any shifted rotated, highly ill-conditioned or extremely complex function. On the other hand, all the functions comprised present a recognizable overall shape and are potentially solvable using limited FEs. For these reasons, Set A is a good environment to test the techniques integrated into AsBeC.

Each function is investigated with 10 dimensions and a maximum number of FEs equal to  $10^3$ . Each experiment is repeated 300 times. This settings fit the main target of the AsBeC algorithm, i.e., finding the solution of low-dimensional problems using few FEs. Besides, a long term setting is also introduced, investigating 30 dimensions with maximum number of FEs equal to  $5 \cdot 10^4$  and each experiment repeated 30 times. This last setting is in line with other ABC works (e.g., Akay and Karaboga, 2012) and it is important to assess premature clustering tendency and refining skills on higher dimensional problems when a lot of FEs are allowed.

**Algorithm 1** The AsBeC pseudo-code

---

```

{Initialization}
1: Generate max(number of seed solutions,  $2 \cdot D + 1$ ) starting random solutions in the search area (Sec. 3.4)
2: Evaluate their quality
3: Identify seed solutions as the best starting solutions and assign employees to them
4: Find minimum of the quadratic model near the best seed solution (Sec. 3.4)
5: if a new best solution has been found then update best seed solution end if
   {SGO phase (Sec. 3.5)}
6: if quasi-separability test passed then use interleaved S.T.E.P solver end if
7: if a new best solution has been found then update best seed solution end if
8: repeat
   {Employees phase}
9:   for all employees do
10:     Generate new pseudo-random solution near its seed solution
11:     Evaluate the quality of solution
12:     if it is better than current employee's seed solution then update seed solution end if
13:   end for
   {Onlookers phase}
14:   Assign onlookers to the seed solutions depending on their quality, according to a biased deterministic rule (Sec. 3.1)
15:   for all onlookers do
16:     for all postponed dance iterations (Sec. 3.2)
17:       Generate new solution near its seed solution, possibly using LI (Sec. 3.3)
18:       Evaluate the quality of solution
19:       if it is better than current onlooker's seed solution then update seed solution end if
20:     end for do
21:   end for
   {Quadratic Prophet phase (Sec. 3.4)}
22:   for all seed solution do
23:     Collect the samplings nearest to the current seed solution
24:     Build quadratic model in its neighbourhood and find its minimum
25:     if it is better than current seed solution then update seed solution end if
26:   end for
   {Scout phase}
27:   if a seed solution is not improved for a limited time then replace it with a new random solution end if
   {Best-so-far}
28:   if a new best solution has been found then update global best end if
29: until requirements are met

```

---

$10^{-16}$  is considered as the minimum achievable by each function of Set A, i.e., it is the selected *tolerance*, in line with the numerical floating point double precision accuracy of MATLAB. The authors of this paper verified that increasing the value of this tolerance does not affect the qualitative interpretation of results by using the performance metrics defined in Section 5. Each run of the algorithm over a function stops at the maximum allowed number of FEs. Repeating many times the experiment helps in reducing the influence of the random component.

#### 4.2 Set B

The expensive optimization session at CEC conference 2015<sup>1</sup> is based on a benchmark, here called Set B, developed by Qu et al. (2016) of 15 functions. The tolerance used is  $10^{-8}$  and the number of repetitions is set to 20. The functions are tested with 10 dimensions and 500 FEs and with 30 dimensions and 1500 FEs.

Set B has been explicitly targeted to costly optimization and therefore it is appropriate for testing AsBeC. Its functions are shifted rotated, thus analytically not separable, and it includes highly ill-conditioned and very

<sup>1</sup> [http://www3.ntu.edu.sg/home/EPNSugan/index\\_files/CEC2015/CEC2015.htm](http://www3.ntu.edu.sg/home/EPNSugan/index_files/CEC2015/CEC2015.htm)

N#	Name	Formulation	Domain	Properties	Minimum
$f_1$	Sphere	$f(x) = \sum_{i=1}^D x_i^2$	$[-100, 100]^D$	U, S, Simple bowl	$f(0, \dots, 0) = 0$
$f_2$	QuarticR	$f(x) = \sum_{i=1}^D i x_i^4 + rnd(0, 1)$	$[-1.28, 1.28]^D$	U, S, Noisy bowl	$f(0, \dots, 0) = 0$
$f_3$	Step	$f(x) = \sum_{i=1}^D (\lfloor x_i + 0.5 \rfloor)^2$	$[-100, 100]^D$	U, S, Jagged bowl	$f(0.5, \dots, 0.5) = 0$
$f_4$	DixonPrice	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^D f(2x_i^2 - x_{i-1})^2$	$[-10, 10]^D$	U, N, Valley	$f(1, \dots, 2 - \frac{2D-2}{2D}) = 0$
$f_5$	Powell	$f(x) = \sum_{i=1}^{D/4} [(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^2 + 10(x_{4i-3} - x_{4i})^4]$	$[-4, 5]^D$	U, N, Valley	$f(3, -1, 0, 1, \dots, 3, -1, 0, 1) = 0$
$f_6$	Rosenbrock	$f(x) = \sum_{i=1}^{D-1} [100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2]$	$[-30, 30]^D$	U, N, Valley	$f(1, \dots, 1) = 0$
$f_7$	Schweffel1.2	$f(x) = \sum_{i=1}^D  \sum_{j=1}^i (x_j) ^2$	$[-100, 100]^D$	U, N, Ridges	$f(0, \dots, 0) = 0$
$f_8$	Schweffel2.22	$f(x) = \sum_{i=1}^D  x_i  + \prod_{i=1}^D  x_i $	$[-10, 10]^D$	U, N, Ridges	$f(0, \dots, 0) = 0$
$f_9$	Zakharov	$f(x) = \sum_{i=1}^D (x_i)^2 + (\sum_{i=1}^D 0.5i(x_i))^2 + (\sum_{i=1}^D 0.5i(x_i))^4$	$[-5, 10]^D$	U, N, Large plate	$f(0, \dots, 0) = 0$
$f_{10}$	Alpine	$f(x) = \sum_{i=1}^D  x_i \sin(x_i) + 0.1x_i $	$[-10, 10]^D$	M, S, Many far minima	$f(0, \dots, 0) = 0$
$f_{11}$	Rastrigin	$f(x) = 10D + \sum_{i=1}^D (x_i^2 - 10 \cos(2\pi x_i))$	$[-5.12, 5.12]^D$	M, S, Several minima	$f(0, \dots, 0) = 0$
$f_{12}$	Ackley	$f(x) = -20 \exp(-0.2 \sqrt{\frac{1}{D} \sum_{i=1}^D x_i^2}) - \exp(\frac{1}{D} \sum_{i=1}^D (\cos(2\pi x_i))) + 20 + \exp(1)$	$[-32, 32]^D$	M, N, Narrow hole	$f(0, \dots, 0) = 0$
$f_{13}$	Griewank	$f(x) = \sum_{i=1}^D \frac{x_i^2}{4000} - \prod_{i=1}^D \cos(\frac{x_i}{\sqrt{i}}) + 1$	$[-600, 600]^D$	M, N, Several minima	$f(0, \dots, 0) = 0$
$f_{14}$	Levy	$f(x) = \sin^2(\pi \omega_1) + \sum_{i=1}^{D-1} (\omega_i - 1)^2 [1 + 10 \sin^2(\pi \omega_i + 1)] + (\omega_D - 1)^2 [1 + \sin^2(2\pi \omega_D)]$ where $\omega_i = 1 + \frac{x_i - 1}{4} \nu_i$	$[-10, 10]^D$	M, S, Many far minima	$f(1, \dots, 1) = 0$
$f_{15}$	Penalized	$f(x) = \frac{\pi}{2} (10 \sin^2(\pi y_1) + (S + (y_D - 1)^2)) + T$ where $y_i = 1 + 0.25(x_i + 1)$ , $S = \sum_{i=1}^{D-1} (y_i - 1)^2 (1 + 10 \sin^2(\pi y_{i+1}))$ and $T = \sum_{i=1}^D [(100(x_i - 10)^4)(x_i > 10) + (100(-x_i - 10)^4)(x_i < -10)]$	$[-50, 50]^D$	M, N, Many far minima	$f(0, \dots, 0) = 0$
$f_{16}$	Penalized2	$f(x) = 0.1(\sin^2(\pi x_i) + (S + ((x(D) - 1)^2)(1 + \sin^2(2\pi x_d)))) + T$ where $S = \sum_{i=1}^{D-1} (x_i - 1)^2 (1 + \sin^2(3\pi x_{i+1}))$ and $T = \sum_{i=1}^D [(100(x_i - 5)^4)(x_i > 5) + (100(-x_i - 5)^4)(x_i < -5)]$	$[-50, 50]^D$	M, N, Many close minima	$f(0, \dots, 0) = 0$
$f_{17}$	Schaffer	$f(x) = 0.5 + \frac{(\sin(\sqrt{\sum_{i=1}^D x_i^2}) - 0.5)}{(1 + 0.001 \sum_{i=1}^D x_i^2)^2}$	$[-100, 100]^D$	M, N, Few far minima	$f(0, \dots, 0) = 0$
$f_{18}$	Whitley	$f(x) = \sum_{i=1}^D \sum_{j=1}^D (((100(x_i^2 - x_j)^2 + (1 - x_j)^2)/4000 - \cos(100(x_i^2 - x_j)^2 + (1 - x_j)^2) + 1)$	$[-10.24, 10.24]^D$	M, N, Many close minima	$f(1, \dots, 1) = 0$

Table 1: The analytical Set A. U=unimodal; M=multimodal; S=separable; N=non-separable;

complex problems, also with extremely jagged shapes. Set B is complementary to Set A, since it tests the adaptability of the optimization methods also to these kinds of functions. Notice that the maximum number of admitted FEs in 30 dimensions is really small ( $50 \cdot D$ ), which is likely to favor a simple exploration of the space. Further details about the settings and functions can be found in the work by Chen et al. (2015).

#### 4.3 Set C

The noiseless Black-Box Optimization Benchmarking (BBOB)<sup>2</sup> set used at CEC 2015 and at GECCO 2015 conferences, here called Set C, is developed by Hansen et al. (2009a and 2009b) and contains 24 functions. The tolerance and the number of repetitions are respectively set to  $10^{-8}$  and to 20. In the present work, the functions are tested with 5 dimensions and 500 FEs and with 20 dimensions and 2000 FEs.

The major advantage of using this benchmark is due to its schematic structure and to the huge amount of literature dealing with it; in fact, at least 151 algorithms have been compared so-far on this benchmark from 2009 to 2015. Set C is not explicitly targeted to costly optimization, since its functions are not meant to be solved within a strictly limited number of FEs. Set C comprises five different groups of functions: (i) separable, (ii) low-conditioned, (iii) high-conditioned, (iv) multi-modal with adequate global structure and (v) multi-modal

<sup>2</sup> <http://coco.gforge.inria.fr>

with weak global structure. This well organized heterogeneous architecture allows to interpret the strengths and weaknesses of each optimization algorithm. For details about this benchmark refer to Hansen et al. (2009b).

## 5 Mean logarithmic value metrics

Some quantitative metrics are introduced to quickly compare algorithms in terms of the objective value reached after a given amount of FEs. A performance estimator for an algorithm is defined as the Logarithmic Value  $LV$ :

$$LV(f, FEs) := \log \left[ \frac{(M_{rep}(f_{best}(f, FEs, rep)) - f_{best}^*(f))}{tolerance} \right]$$

where  $rep$  is the run repetition,  $FEs$  is the number of performed function evaluations,  $f$  is the selected benchmark function,  $f_{best}^*$  is the analytical optimal value,  $f_{best}$  is the best objective value obtained and  $M_{rep}$  represents the median operator over repetitions. The median is chosen in spite of the mean since it is less sensible to outliers. The numerator represents the *residual* and the  $LV$  represents the median performance of an algorithm in terms of orders of magnitude missing to the given *tolerance*. A  $LV$  is always nonnegative and it reaches zero when the median results of an algorithm are optimal, within the given tolerance.

Three averaged forms of the Logarithmic Value, called Mean Logarithmic Values ( $MLVs$ ), are defined.  $MLV_{FEs}$  is an average over the number of benchmark functions  $N_f$ ,  $MLV_f$  is an average over the maximum number of function evaluations  $FEs_{max}$  and  $MLV_A$  is an average over both functions and FEs:

$$MLV_{FEs}(FEs) = \frac{1}{N_f} \cdot \sum_{f=1}^{N_f} LV(f, FEs)$$

$$MLV_f(f) = \frac{1}{FEs_{max}} \cdot \sum_{FEs=1}^{FEs_{max}} LV(f, FEs)$$

$$MLV_A = \frac{1}{FEs_{max}} \cdot \sum_{FEs=1}^{FEs_{max}} \frac{1}{N_f} \sum_{f=1}^{N_f} LV(f, FEs)$$

$MLV_A$  helps in comparing different algorithms through just one real number. It is meaningful since it considers the performance evolution on the whole set of functions, not only after a specific number of FEs.

The main difference of the introduced metric with respect to a simple mean of the median residuals is the fact that  $MLVs$  average the logarithms of the residuals, and not the mean of the actual values. This helps to assess how much the algorithm is getting closer to the optimum in relative terms.

The Logarithmic Value and its averaged versions are here introduced for the first time by the authors of this paper. They will be used for comparisons alongside other classical and benchmark-specific metrics.

## 6 Analysis of the improving techniques on Set A

Having defined the comparison metrics, the main three parameters of the ABC architecture, i.e., the colony size, the number of cycles and the limit parameter, are chosen. The limit parameter is set on  $D \cdot SN$  as advised for the original ABC (Karaboga and Akay, 2007), while the number of cycles depends on FEs and on the colony size. The overall number of agents,  $N$ , is a key performance-driving parameter for population-based algorithms. Six configurations for the total number of individuals (4, 8, 16, 32, 64 and 128 individuals) are directly investigated and  $MLV_A$  is the metric used to choose the best one. For the original ABC the best colony size turns out to be 8 (see Table 2). This properly tuned ABC is used for the comparisons.

At this point, the impact of each single technique proposed in Section 3 is evaluated. They are applied one-by-one to the reference ABC configuration. Fig. 1a illustrates the correspondent  $MLV_A$  for each technique and for the basic ABC. All the techniques have positive effects on the algorithm performance. The QP and SGO clearly take a dominant role with respect to the other techniques, that show a more limited impact.

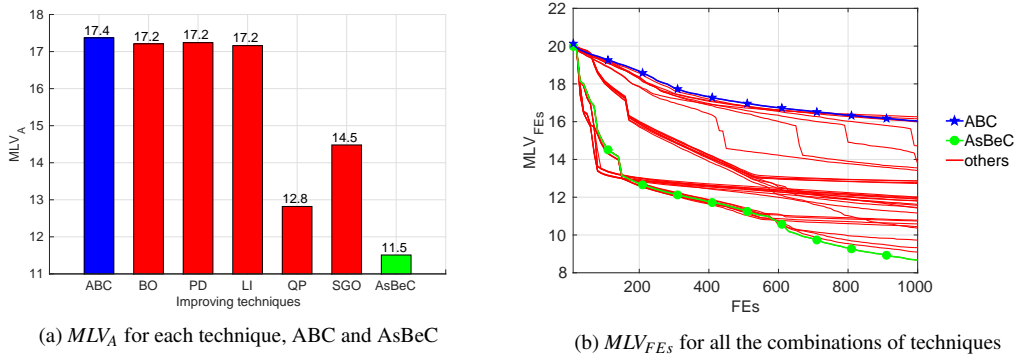


Fig. 1: Validation of the techniques using the  $MLV$  metrics

The differences between the *ABC* with or without the implemented techniques presented in Fig. 1a are statistically significant. The  $MLV_A$  represents the average of the  $LV$  over the evaluations and the functions. A Student significance t-test has been conducted over the two samples (the samples mean will follow a normal distribution by the Central Limit Theorem). The null hypothesis that the  $MLV_A$  of the modified algorithm is greater than the one of *ABC* has been tested against the alternative hypothesis that the mean has positive value. In all the cases, the null hypotheses can be rejected with significance level of 0.05.

Then, all possible combinations of techniques are tested in order to capture their mutual interactions. The  $MLV_{FEs}$  comparisons are shown in Fig. 1b, where each line corresponds to a different combination of techniques. The best configuration (AsBeC line) is the one with lowest  $MLV_A$ , and it corresponds to the AsBeC ( $MLV_A$  of 11.5, reported in Fig. 1a). The Quadratic Prophet, based on a trust region approach, is revealed to be impressively suited for this kind of problems, as expected.

Once the benefit of the techniques is clear and AsBeC is recognized as the best combination of these techniques, the number of agents is studied for Set A. The best number of bees results 8 for the setting with 10 dimensions and 32 for the setting with 30 dimensions (see Table 2). Indeed, other tests on different dimensions showed that the number of individuals should be at least equal to 8 and approximately proportional to the number of dimensions. The same AsBeC settings discussed in this section for set A will be also used for set B and C, in order to prove the robustness of this algorithm configuration.

## 7 Comparison with other algorithms on Set A

The AsBeC effectiveness has to be tested when compared to other state-of-the-art methods of the same class suitable for the same goal (see Section 1), consistently with what has been done by Akay and Karaboga (2012) or Gao and Liu (2011) for their algorithms. According to the considerations in Section 2, GABC, JA-ABC5 and RABC are selected as representative of ABC modifications also suited for the specific goal under study. The authors have selected other three typologies of direct search algorithms for the comparison: FIREFLY (Yang, 2009), CMA-ES (Hansen, 2006) and BOBYQA (Powell, 2009). FIREFLY is a notable competitor of the ABC among the swarm-based techniques, while the evolutionary CMA-ES and the trust region BOBYQA are recognized as state-of-the-art in their fields and have been widely used for engineering applications.

The implementations used for FIREFLY and CMA-ES are the ones made available by the authors themselves. The MATLAB routine used for BOBYQA is instead the one embedded in the NAG optimization toolbox<sup>3</sup>.

<sup>3</sup> <http://www.nag.co.uk/numeric/MB/start.asp>

### 7.1 Choice of parameter values on Set A

A key point when comparing different methods consists in tuning the parameters of the optimizers to make them perform at their best. For all the algorithms considered, the parameter selection is addressed as in Section 6 for ABC and AsBeC. Only the number  $N$  of individuals for population-based methods and the number  $N$  of interpolation samples for BOBYQA are directly investigated on Set A. The range studied for  $N$  is a progression of the power of two; its boundaries depend on the specific mechanics driving each algorithm. The best size is selected according to the  $MLV_A$  metric and is reported in Table 2.

All other parameters assume the values suggested in literature as suitable for many problems (Zhu and Kwong, 2010 for GABC, Kang et al., 2011 for RABC, Sulaiman et al., 2015 for JA-ABC5, Hansen, 2006 for CMA-ES and Powell, 2009 for BOBYQA). For FIREFLY, the authors use the parameters proposed by Yang (2010) for the MATLAB implementation of the algorithm, since they reach better average results on the adopted benchmark with respect to the general suggestions by Yang and He (2013).

Dimensions	ABC	GABC	RABC	JA-ABC5	FIREFLY	CMA-ES	BOBYQA	AsBeC
10D	8	8	8	16	16	8	32	8
30D	16	16	16	16	32	32	256	32

Table 2: Best  $N$  for each algorithm according to the  $MLV_A$  metric

### 7.2 Results of comparison on Set A

The analysis of  $MLV_{FEs}$  for 10 dimensions is reported in Figure 2a. It shows that AsBeC is the most effective on average for quickly improving the solution as well as refining it. Referring to function-by-function performance in the electronic appendix, the AsBeC algorithm is able to outperform, or at least to approach, all the others during the entire FE envelope, on all the benchmark functions. In short, its performance is very good and robust, as it is revealed also by the boxplot comparisons in the electronic appendix. None of the algorithms appears remarkably better than AsBeC in any function. AsBeC performs always better than the original ABC and it is the only algorithm able to get close to the global optimum of the Shaffer function. Final objective values analysis, reported in the electronic appendix, recognize AsBeC as the most appropriate for the benchmark. Moreover, following the same procedure adopted in Section 6, the improvements of the  $MLV_A$  of AsBeC with respect to the other algorithms show statistical significance.

Considering the other algorithms, the two variants GABC and JA-ABC5 perform slightly better than the original ABC. Instead, RABC is in general faster and more capable also at refining the solution. Nonetheless,

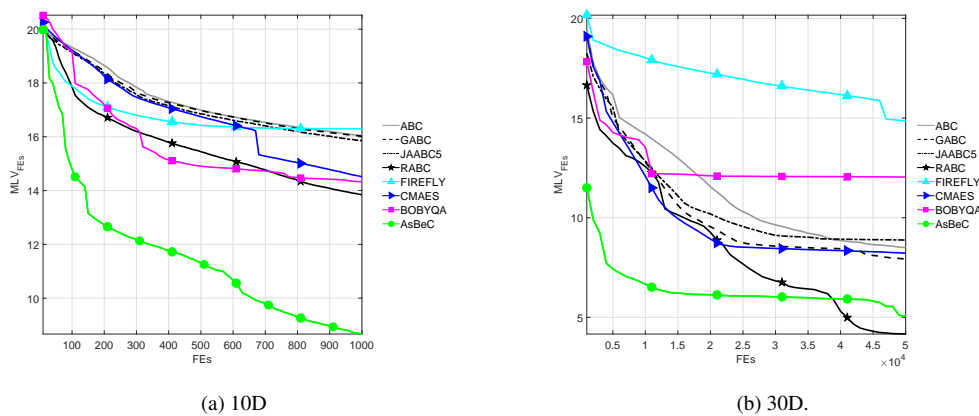


Fig. 2: Set A. Comparison of the algorithms using  $MLV_{FEs}$

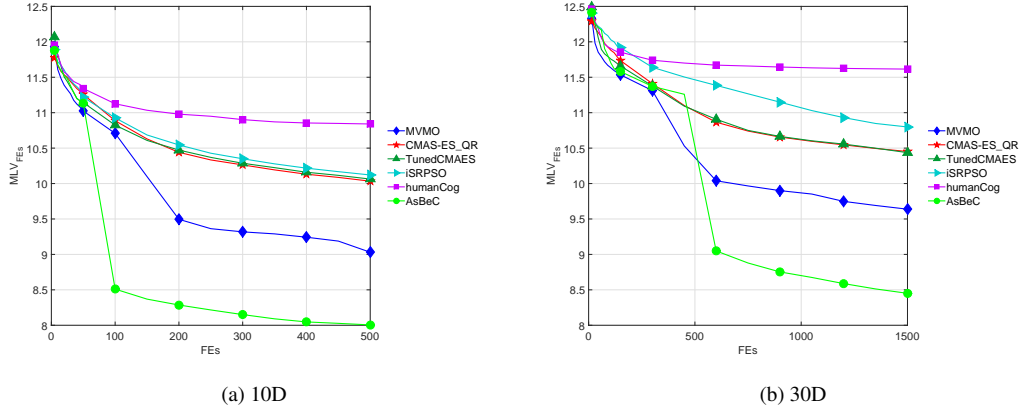


Fig. 3: Set B. Comparison of the algorithms using  $MLV_{FEs}$

it is worse than the original ABC in those functions where finding a preferential direction towards the global optimum is particularly difficult. FIREFLY is among the fastest ones in reaching the best regions, but its refining ability is poor. The CMA-ES performance is robust and shows a good growth potential, but it needs more total FEs to improve. BOBYQA presents the most fluctuating results, performing very well on some functions but getting stuck in local minima in many others.

Looking at the overall results, it is clear that they are in general not convergent and in many functions they are still far from the global minimum. Nevertheless, the above comparison is still solid. In fact, in real-like problems where the optimal position is unknown, designers are interested in obtaining the best solution improvement within a limited amount of resources.

Set A with a 30-dimensions setting is not the main focus of the paper, but provides interesting insights on long-term performance. The analysis of the MLV in Figure 2b and function-by-function performance reported in the electronic appendix, reveal that AsBeC is the most promising for quickly improving the solution also in this kind of problems. In Figure 2b, first data are plotted after  $10^3$  FEs to improve visualization, showing that AsBeC reaches a very low value after very few FEs. Moreover, AsBeC is able to reach the given tolerance in 11/18 of the functions and, among the other functions, only Dixon Price and Rosenbrock can be solved by other algorithms. In short, AsBeC is the best algorithm up to around  $4 \cdot 10^4$  FEs. RABC is the most efficient at refining if a large number of FEs is available. However, it is clearly much slower than AsBeC. The other two ABC variants and CMA-ES are quite similar and better than ABC, as revealed by the  $MLV_{FEs}$ . BOBYQA is again among the fastest ones during the very first optimization phases. FIREFLY seems to show a slower convergence speed with respect to the others when dimensionality rises.

## 8 Other validations

The primary goal of the two validation sets B and C is to assess the robustness of AsBeC, besides its quality, in comparison to the most recent and proficient optimization methods. For this reason, the same AsBeC standard settings discussed in Section 6 for Set A will be also used for Set B and C. Following the guidelines of Section 7.1, the bee colony size is chosen as  $N = 8$  for  $D = 5$  and  $D = 10$ ,  $N = 20$  for  $D = 20$ , and  $N = 32$  for  $D = 30$ .

### 8.1 Results of the comparison on Set B

Public results on this benchmark<sup>1</sup> by Qu et al. (2016) include the following algorithms:

- MVMO, a population-based stochastic technique with a mapping function for the offspring;

- CMAS-ES QR and TunedCMAES, variants of CMAES for expensive scenarios;
- iSRPSO, a PSO implementing a dynamic learning strategy for velocity updating;
- humanCog, a 3-layer architecture that mimics human cognitive behaviour.

The  $MLV_{FEs}$  comparisons with the AsBeC are presented in Figure 3. Extended function-by-function plots of median residual over repetitions as function of FEs and complete tables with mean, median and standard deviation for final achievements are reported in the electronic appendix. It is evident that AsBeC has a great overall quality, reaching much lower values than the best method, MVMO. AsBeC is able to solve the quadratic problems, even if ill-conditioned, to recognize the quasi-separability of the test function 4, and to approach the best methods in the other functions. Unlike the other algorithms, AsBeC is capable of a stable performance regardless the dimensionality.

Besides, Qu et al. (2016) defined an official total scoring for this competition as an average of the median and mean values at the end and at the middle of the computation. The AsBeC total score is one order of magnitude less than MVMO (full table reported in the electronic appendix). Notice that this scoring favors the improvements made on the most difficult function, TF 10. Since both AsBeC and MVMO are able to better minimize this problem, their scores are much lower than those of the other algorithms.

## 8.2 Results of comparison on Set C

At the CEC and GECCO conferences in 2015, 26 algorithms have competed on this benchmark<sup>2</sup>. Indeed, they are tuned variants of few basic methods. Among them, four top categories can be identified:

- CMAES derived, including many tuned variants of IPOP-CMAES;
- Surrogate based, including the ones that exploit the MATLAB MATSuMoTo Library for metamodels;
- DE derived, tuned for cheap, medium, and expensive settings;
- Axis-Parallel Brent-S.T.E.P. method, which investigates some variants of the multidimensional Brent-S.T.E.P. method.

For each category, the best algorithm in terms of performance and robustness according to the official results of the conference is compared to the standard AsBeC through the  $MLV_{FEs}$  metric. These four tuned methods are the GP5-CMAES, RAND- 2xDefaultMATSuMoTo, R-DE-10e2 and Srr.

The  $MLV_{FEs}$  metric in Figure 3 recognizes the Srr algorithm as the best one, just slightly better than AsBeC. However, function-by-function plots and BBOB official ECDF metric, reported in the electronic appendix, point out that Srr is less robust, especially when addressing non separable multi-modal functions. Srr is the best method for separable problems (f1-5), but it is the worst on some others, reaching premature convergence (e.g.,

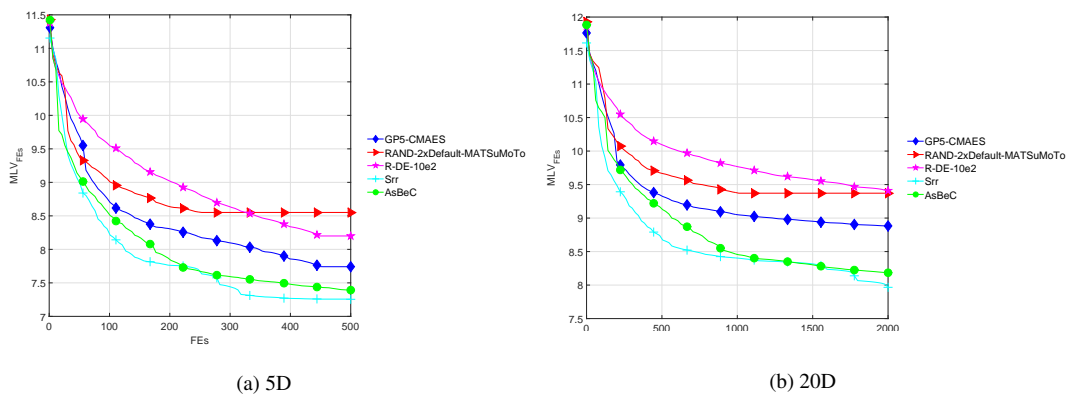


Fig. 4: Set C. Comparison of the algorithms using  $MLV_{FEs}$



in f7 and f18). Instead, AsBeC and GP5-CMAES offer overall stable high-quality performance. In particular, AsBeC reaches very good results when solving separable, low-conditioned and weak multi-modal problems.

In all the functions, AsBeC is at least close to the best methods, even if no specific tuning has been studied for this set.

## 9 Concluding remarks

Achieving fast and robust improvements in single-objective optimization problems involving expensive analyses means saving precious time and resources. A new swarm-based algorithm hybridized with interpolation strategies, called Artificial super-Bee enhanced Colony (AsBeC) algorithm, is proposed. The new algorithm has been designed for solving expensive problems with low dimensionality, using a limited number of function evaluations. The ambition of this work is to improve the local search, i.e., the exploitation ability, concurrently preserving the good global attitude of the original ABC, i.e., exploration ability, especially during the first search phases.

A meaningful metric for comparison is defined, the Mean Logarithmic Value, which takes into account the evolution of the optimization process and the relative distance to the analytical optimum. All the implemented techniques are analysed in order to identify a standard robust setting for them. The standard AsBeC algorithm is compared with ABC, with some of its relevant modifications and with other state-of-the-art direct search algorithms. The same standard AsBeC has also been validated on other benchmarks, and it is compared with some of the latest methods presented at the CEC 2015 and GECCO 2015 conferences, which have specifically been tuned for the benchmarks. The AsBeC algorithm is confirmed to be robust and effective on all the benchmarks.

This promising outcome paves the way for a useful application of the proposed algorithm, especially in engineering. Indeed, the basic principles presented in this paper already showed interesting results in the past when applied by the authors to turbine design (Bertini et al., 2013). Future works will include comparisons on real-world optimization problems, introduction of parallel strategies and extension to multi-objective problems.

**Acknowledgements** The authors would like to thank GE Avio S.r.l. and its Engineering Technologies department, especially Ing. F. Bertini and Ing. E. Spano. Their collaboration was fundamental for shaping the AsBeC algorithm within an industrial application framework. We are also grateful to Prof. E. Benini (Università di Padova) for his comments and reviews.

## References

- Akay, B., & Karaboga, D. (2010). Artificial bee colony algorithm for large-scale problems and engineering design optimization. *Journal of Intelligent Manufacturing*, 23(4), 1001-1014.
- Akay, B., & Karaboga, D. (2012). A modified artificial bee colony algorithm for real-parameter optimization. *Information Sciences*, 192(0), 120-142.
- Aydin, D. (2015). Composite artificial bee colony algorithms: From component-based analysis to high-performing algorithms. *Applied Soft Computing*, 32(0), 266-285.
- Bagirov, A. M., Karaszen, B. & Sezer, M. (2008). Discrete gradient method: Derivative-free method for nonsmooth optimization. *Journal of Optimization Theory and Application*, 137(2), 317-334.
- Bartz-Beielstein, T., Lasarczyk, C., & Preuß, M. (2005). Sequential parameter optimization. In *Proceedings of the 2005 IEEE Congress on Evolutionary Computation* (Vol. 1, pp. 773-780). Piscataway, NJ: IEEE Press.
- Baudiš, P., & Pošík, P. (2015). Global Line Search Algorithm Hybridized with Quadratic Interpolation and Its Extension to Separable Functions. In *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation* (pp. 257-264). New York, NY: ACM.
- Berghen, F. V., & Bersini, H. (2005). CONDOR, a new parallel, constrained extension of Powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. *Journal of Computational and Applied Mathematics*, 181(1), 157-175.

- Bertini, F., Dal Mas, L., Vassio, L., & Ampellio, E. (2013). Multidisciplinary optimization for gas turbines design. XXII AIDAA Conference. <http://arxiv.org/pdf/1402.0420v1.pdf>. Accessed February 5, 2016.
- Beyer, H.G. & Schwefel, H.P. (2002). Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1), 3-52.
- Bolaji, A., Khader, A., Al-Betar, M., & Awadallah, M. (2013). Artificial Bee Colony algorithm, its variants and applications: a survey. *Journal of Theoretical and Applied Information Technology*, 47(2), 434-459.
- Box, G. E. P., & Draper, N. (2007). *Response surfaces, mixtures, and ridge analyses*, Second Edition. Hoboken, NJ: John Wiley & Sons.
- Buhmann, M. D. (2003). *Radial Basis Functions: Theory and Implementations*. Cambridge, UK: Cambridge University Press.
- Chen, Q., Liu, B., Zhang, Q., Liang, J. J., Suganthan, P. N. & Qu B. Y (2015). Problem Definitions and Evaluation Criteria for CEC 2015 Special Session on Bound Constrained Single-Objective Computationally Expensive Numerical Optimization. Technical Report, Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou, China, Nov 2014.
- Conn, A. R., Gould, G. & Toint, P. (2000). *Trust region methods*. Philadelphia, PA: SIAM.
- Couzin, I. D., Krause, L., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 128(1), 1-11.
- Deb, K. (2012). *Optimization for engineering design: Algorithms and examples*, Second Edition. New Delhi, India: PHI Learning Pvt.
- Dorigo, M., & Stützle, T. (2004). *Ant colony optimization*. Cambridge, MA: MIT Press.
- El-Abd, M. (2011). Opposition-based artificial bee colony algorithm. In *Proceedings of the 13th annual Conference on Genetic and Evolutionary Computation* (pp. 109-116). New York, NY: ACM.
- Floudas, C., & Pardalos, M. (2009). *Encyclopedia of Optimization*, Second Edition. New York, NY: Springer.
- Frish, K. (1967). *The Dance Language and Orientation of Bees*. Cambridge, MA: Harvard University Press.
- Gao, W., & Liu, S. (2011). Improved artificial bee colony algorithm for global optimization. *Information Processing Letters*, 111, 871-882.
- Gao, W., & Liu, S. (2012). A global best artificial bee colony algorithm for global optimization. *Journal of Computational and Applied Mathematics*, 236(11), 2741-2753.
- Goldberg, D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA: Addison-Wesley Longman Publishing.
- Hansen, N. (2006). The CMA Evolution Strategy: A Comparing Review. In Lozano et al., *Towards a new evolutionary computation. Advances in estimation of distribution algorithms* (pp. 75-102). Berlin, Germany: Springer.
- Hansen, N., Auger, A., Finck, S., & Ros, R. (2009a). Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA.
- Hansen, N., Finck, S., Ros, R., & Auger, A. (2009b). Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA.
- Iliadis, J., & Jayne, C. (Eds.). (2015). *Engineering Applications of Neural Networks: Proceedings of the 16th International Conference on Engineering Applications of Neural Networks*. Berlin, Germany: Springer.
- Jamil, M., & Yang, X. S. (2013). A literature survey of benchmark functions for global optimization problems. *International Journal of Mathematical Modelling and Numerical Optimization*, 4(2), 150-194.
- Jones, D. R., Schonlau, M., & Welch W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4), 455-492.
- Kang, F., Li, J., & Xu, Q. (2009). Structural inverse analysis by hybrid simplex artificial bee colony algorithms. *Computers and Structures*, 87(13-14), 861-870.
- Kang, F., Li, J., & Ma, Z. (2011). Rosenbrock artificial bee colony algorithm for accurate global optimization of numerical functions. *Information Sciences*, 181(16), 3508-3531.
- Kang, F., Li, J. & Li, H. (2013). Artificial bee colony algorithm and pattern search hybridized for global optimization. *Applied Soft Computing*, 13(4), 1781-1791.

- Karaboga, D. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of Global Optimization*, 39(3), 459-471.
- Karaboga, D., & Akay, B. (2007). Artificial Bee Colony (ABC) optimization algorithm for solving constrained optimization problems. In Melin et al., *Foundations of Fuzzy Logic and Soft Computing: Proceedings of the 12th International Fuzzy Systems Association World Congress 2007* (pp. 789-798). Berlin, Germany: Springer.
- Karaboga, D., & Akay, B. (2009). A comparative study of Artificial Bee Colony Algorithm. *Applied Mathematics and Computations*, 214(1), 108-132.
- Karaboga, D., Gorkemli, B., Ozturk, C., & Karaboga, N. (2014). A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, 42(1), 21-57.
- Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. In *Proceedings of IEEE International Conference on Neural Networks 1995* (pp. 1942-1948). Piscataway, NJ: IEEE Press.
- Kern, S., Hansen, N., & Koumoutsakos, P. (2006). Local meta-models for optimization using evolution strategies. In Runarsson et al., *Proceedings of the 9th International Conference of Parallel Problem Solving from Nature - PPSN IX* (pp. 939-948). Berlin, Germany: Springer.
- Kong, X., Liu, S., & Wang, Z. (2013). An improved Artificial Bee Colony algorithm and its application. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 6(6), 259-274.
- Koziel, S., & Leifsson, L. (2013). *Surrogate-Based Modeling and Optimization, Applications in Engineering*. New York, NY: Springer.
- Lagaros, N., & Papadrakakis, M. (Eds.). (2015). *Computational Methods in Applied Sciences: Engineering and Applied Sciences Optimization*. Cham, Switzerland: Springer.
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least square. *Quarterly Journal of Applied Mathematics*, 2, 164-168.
- Li, B., & Li, Y. (2010). BE-ABC: Hybrid Artificial Bee Colony Algorithm with Balancing Evolution Strategy. In *Proceedings of the 3rd International Conference on Intelligent Control and Information Processing* (pp. 217-222). Piscataway, NJ: IEEE Press.
- Liao, T., Aydin, D., & Stützle, T. (2013). Artificial bee colonies for continuous optimization: Experimental analysis and improvements. *Swarm Intelligence*, 7(4): 327-357.
- Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11(2), 431-441.
- Powell, M. (2000). UOBYQA: Unconstrained optimization by quadratic approximation. Technical Report No. DAMTP2000/14, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England.
- Powell, M., (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical report NA2009/06, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England.
- Price, K., Storn, R., & Lampinen, J. (2005). *Differential Evolution - A Practical Approach to Global Optimization*. Berlin, Germany: Springer.
- Qu, B. Y., Liang, J. J. , Wang, Z. Y. , Chen, Q., & Suganthan, P. N. (2016). Novel Benchmark Functions for Continuous Multimodal Optimization with Comparative Results. *Swarm and Evolutionary Computation*, 26, 23-34.
- Rao, S.(2009). *Engineering Optimization: Theory and Practice*, Fourth Edition. New York, NY: Wiley.
- Rios, L. M., & Sahinidis, N. V., (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3), 1247-1293.
- Roland, F. W., & Nachtigal, N. M. (1991). QMR: A quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik*. 60(1), 315-339.
- Rosenbrock, H. H., (1960). An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3), 175-184.
- Simpson, T. W., Mauery, T. M., Korte, J. J., & Mistree, F. (2001). Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA Journal*, 39(12), 2233-2241.

- Subotic, M., Tuba, M., & Stanarevic, N. (2011). Different approaches in parallelization of the artificial bee colony algorithm. *International Journal of Mathematical Models and Methods in Applied Sciences*, 5(4), 755-762.
- Subotic, M. (2011). Artificial bee colony algorithm with multiple onlookers for constrained optimization problems. In *Proceedings of the 5th European Computing Conference* (pp. 251-256). Stevens Point, WI: World Scientific and Engineering Academy and Society.
- Sulaiman, N., Mohamed-Saleh, J., & Abro, A., G. (2014). New enhanced Artificial Bee Colony (JA-ABC5) algorithm with application for reactive power optimization. *The Scientific World Journal*, 2015, 1-11.
- Swartzberg, S., Seront, G., & Bersini, H. (1994). S.T.E.P.: the easiest way to optimize a function. In *Proceedings of the 1st IEEE Conference on Evolutionary Computation* (pp. 519-524). Piscataway, NJ: IEEE Press.
- Talbi, E. (2009). *Metaheuristics: from design to implementation*. Hoboken, NJ: Wiley.
- Yang, X. (2009). Firefly algorithms for multimodal optimization. In *Proceedings of the 5th International Conference on Stochastic Algorithms: Foundations and Applications* (pp. 169-178). Berlin, Germany: Springer.
- Yang, X. (2010). *Engineering Optimization: An Introduction with Metaheuristic Applications*. Hoboken, NJ: Wiley.
- Yang, X., & He, X. (2013). Firefly algorithm: recent advances and applications. *International Journal of Swarm Intelligence*, 1(1), 36-50.
- Zhao, H., Pei, Z., Jiang, J., Guan, R., Wang, C., & Shi, X. (2010). A hybrid swarm intelligent method based on genetic algorithm and artificial bee colony. In *Proceedings of the 1st International Conference on Advances in Swarm Intelligence* (pp. 558-565). Berlin, Germany: Springer.
- Zhu, G., & Kwong, S. (2010). Gbest-guided artificial bee colony algorithm for numerical function optimization. *Applied Mathematics and Computation*, 217(7), 3166-3173.