Landscape Features for Computationally Expensive Evaluation Functions: Revisiting the Problem of Noise

Eric O. Scott[(✉)] and Kenneth A. De Jong

Department of Computer Science, George Mason University, Fairfax, VA, USA
{escott8,kdejong}@gmu.edu

Abstract. When combined with machine learning, the black-box analysis of fitness landscapes promises to provide us with easy-to-compute features that can be used to select and configure an algorithm that is well-suited to the task at hand. As applications that involve computationally expensive, stochastic simulations become increasingly relevant in practice, however, there is a need for landscape features that are both (A) possible to estimate with a very limited budget of fitness evaluations, and (B) accurate in the presence of small to moderate amounts of noise. We show via a small set of relatively inexpensive landscape features based on hill-climbing methods that these two goals are in tension with each other: cheap features are sometimes extremely sensitive to even very small amounts of noise. We propose that features whose values are calculated using population-based search methods may provide a path forward in developing landscape analysis tools that are both inexpensive and robust to noise.

Keywords: Parameter tuning · Landscape analysis · Meta-learning · Noisy evaluation

1 Introduction

Tuning the parameters of large, stochastic simulations in science and engineering is becoming an increasingly important and popular application domain for evolutionary algorithms (EAs) and metaheuristics (ex. [6,16,17]). These applications tend to involve fitness functions that are very expensive to compute—each evaluation taking on the order of seconds, minutes, or even hours to complete. To approach problems of this kind effectively, the algorithm designer must have some means of quickly and efficiently gathering information about the problem that can help reduce the number of generations that are necessary for a search method to reach a satisfactory solution. In applications where a thorough analytical understanding of the problem is not available, this information-gathering process is often restricted to learning about the problem by directly sampling the evaluation function, which is treated as a *black box.*

Finding ways of characterizing salient properties of fitness landscapes via empirical data—and, especially, of predicting what kinds of algorithms are likely to perform well on them—has been a fundamental goal of metaheuristics and evolutionary computation research since the early days of the field. Researchers have leveraged a number of different mathematical ideas over the years (such as epistasis, correlational properties, and information theory) to produce several families of black-box landscape features [11,14,18,19]. Because these statistical methods are based solely on queries made to the objective function, they can be used even on poorly-understood problems, where little or nothing is known *a priori* about the relationships among variables.

In order for black-box landscape analysis to be useful in practice, however, the information it provides about how to solve a given problem must outweigh the cost of calculating the statistical features. The 'budget' of computational effort that can be spared for up-front analysis is especially small in applications whose evaluation functions involve expensive scientific simulations. A number of landscape features have been proposed that can be computed effectively with especially few queries to the evaluation function, at least on deterministic (noiseless) test functions [1]. Real-world fitness landscapes, and stochastic simulations in particular, often display some degree of noise, however.

In this paper, we are concerned about the intersection of noisy fitness landscapes and the calculation of informative landscape features for computationally intensive applications. In some circumstances, noise may interfere, not only with the progress of a search algorithm as it seeks a global optimum, but also with the attempts of a landscape analysis tool to accurately estimate properties of the task. The problem of noisy fitness functions was heavily studied in the 1990's and early 2000's, and a variety of well-understood approaches are available for configuring evolutionary algorithms to cope with noise [3,10]. Coping with noise does not come for free, however—it often requires extra fitness evaluations which we may not be able to afford when the evaluation function is computationally intensive.

We find it necessary, then, to revisit the well-studied question of noise, now in the context of a pressing need for effective landscape analysis tools that make as few queries as possible to the evaluation function. In this study, we examine several cheap-to-evaluate landscape features and show that a subset of them are extremely sensitive to even very small amounts of noise. Furthermore, we find that the error that this noise introduces into feature estimation can be difficult to correct for in an efficient way. As an alternative, we propose features that use population-based methods as a means of gathering information about the landscape in a way that is both inexpensive and robust to noise.

1.1 Research Questions

Intuitively, it's clear that qualitative features of an objective function such as multimodality, deceptiveness, or the correlation of traits among parents and offspring [2,13] convey a great deal of information about whether a given search strategy is well-suited to particular task. Early work on landscape analysis sought

to identify ways in which a problem might be "easy" or "hard" for a particular algorithm of choice (namely the genetic algorithm, ex. [8,9]). But as the philosophy of the research community moves toward "solving the problem at hand in the best way possible, rather than promoting a certain metaheuristic" [5], the primary purpose of landscape analysis has shifted to serving as a predictive aid in the design or selection of a custom algorithm that is well-suited to the given task [15]. Landscape features can be used as input data for machine learning algorithms, which are increasingly being used to learn predictive models for use in algorithm selection and configuration (ex. [4]). Even if a particular statistical feature is difficult for an engineer to interpret in terms of intuitive concepts like multimodality, the feature may be useful if it provides salient or complementary information to a machine learner in conjunction with other features.

If there is a great deal of error or bias in an estimate of a feature, however, its usefulness as a basis for learning may in some cases be greatly diminished. There is a practical need, then, for landscape features that are both (A) inexpensive to estimate, and (B) accurate in the presence of small to moderate amounts of noise. Table 1 details a number of features, taken from Abell et al., that can typically be computed in on the order of a few hundred or a few thousand fitness evaluations, but which are still sufficiently informative to enable a portfolio method to perform well on a suite of noiseless benchmark functions [1]. These satisfy our criterion of inexpensiveness (A), but how do they fair with noise (B)?

> **Research Question 1**: How sensitive to noise are the 8 landscape features identified in Table 1?

Next we begin an investigation into how error in the estimation of features can be corrected for. A straightforward way to do this is to seek to approximate the features of the *expected fitness landscape* $\hat{F}(\boldsymbol{x})$ by taking several fitness samples each time the landscape is queried and returning their 'explicit average' [10].

> **Research Question 2**: Is using explicit averaging an effective means of correcting for noise when measuring these features?

Finally, the features in Table 1 rely heavily on the results of a number of runs of a hill-climbing method as a means of exploring the structure of the landscape. Trajectory methods such as this are notorious for their sensitivity to noise. We consider the possibility that a population-based method may be more effective at identifying informative local optima in the presence of noise:

> **Research Question 3**: Can population-based algorithms serve as a useful alternative to hill-climbers for quickly gathering information about noisy fitness landscapes?

Table 1. Landscape features used this study.

	Feature	Description
1	`MeanPairwiseLocalOptDist`	Mean pairwise distance between optima found by a number of hill climbers.
2	`StdPairwiseLocalOptDist`	Standard deviation of (1).
3	`MeanLocalToBestDist`	Mean distance between the best known optimum and the optima found by a number of hill climbers.
4	`StdLocalToBestDist`	Standard deviation of (3).
5	`FractionBest`	The ratio of local optima found by the hill-climbers that have fitness equal to the best known optimum.
6	`MeanRandomToLocalDist`	Mean distance from a number of random points to the nearest optimum found with a hill climber.
7	`StdRandomToLocalDist`	Standard deviation of (6).
8	`FDC`	Local fitness distance correlation, based on the best result of the hill-climbers

2   Methodology

2.1   Test Functions

Our experiments are conducted on 10-dimensional instances from the suite of 24 test functions that are implemented in version v15.03 of the COmparing Continuous Optimisers (COCO) platform, a framework that has been used for a number of years in the Black-Box-Optimization-Benchmarking (BBOB) workshops held at GECCO and CEC. This test suite includes many well-known unimodal and multimodal real-valued functions, such as the sphere, Rastrigin, and Rosenbrock functions, along with rotated variants, etc., all of which are defined on a range of $[-5, 5]$ in each dimension. The COCO source code is available from http://coco.gforge.inria.fr/.

In addressing **RQ1**, our independent variable will be the amount of noise on the landscape. We opt to use a multiplicative noise model of the form

$$F(\boldsymbol{x}) = f(\boldsymbol{x}) + p \cdot |f(\boldsymbol{x}) - f(\boldsymbol{x}^*)| \cdot \epsilon, \tag{1}$$

where $f(\boldsymbol{x})$ is the original (noiseless) test function, $f(\boldsymbol{x}^*)$ is the fitness of the global optimum, and $\epsilon \sim \mathcal{N}(0, 1)$ is a standard Gaussian random variable. The constant p controls the strength of the noise. In this model, the amount of noise that is added to the landscape at the point $\boldsymbol{x}$ is proportional to the difference between its fitness and the global best fitness—so, the poorer a solution is, the nosier it is. This qualitative rule holds in many applications, where poor solutions often correspond to solutions that have especially unstable behavior.

2.2   Features

All of the features in Table 1 make some use of the result of a number of independent runs of a hill-climbing algorithm. We implement the hill climber as a $(1+1)$-style evolutionary algorithm, and we run this method 100 times to gather a set of representative local optima from which features may be computed. Each individual in the EA is represented as a point $\boldsymbol{x} \in \mathbb{R}^L$, with $L = 10$, and we apply a 1-dimensional Gaussian mutation operator to each element of the offspring with probability $1/L$. We let each hill climber run for 3,000 steps, so as to get a stable estimate of the features. It is worth noting, however, that features based on hill climbing can be informative even if they are run only for a very small number of steps [1].

Features 1–4 are computed directly from the best individuals found by the 100 runs. Feature 5 (`FractionBest`) denotes the fraction of the 100 hill-climbing runs whose best individual has fitness equal to the overall best individual found in all 100 runs. The intent of this feature is to measure the frequency with which a greedy search method converges on a local optimum. There is always some variation, however, in just how closely a given climber will converge to the true local optimum. For the purposes of calculating this feature, then, we consider two individuals to have 'equal' fitness if and only if the difference between their fitnesses is less then an arbitrary threshold value of 0.01.

We compute features 6 and 7 using 1,000 random points. For feature 8, we use a local variant of Jones' well-known fitness distance correlation (FDC) [11]. Classical fitness distance correlation requires knowledge of the global optimum to be computed. Since we are using synthetic test functions, we do have knowledge of the global optimum. The purpose of this study, however, is to examine the behavior of landscape features as exploratory, black-box analysis tools. We follow Kallel and Scipemaier in defining a local FDC simply by substituting the best *known* optimum for the global optimum [12]. In our case, the "best known optimum" refers to the best optimum found by the 100 hill-climbing routines.

2.3   Coping with Noise

A test of **RQ2** involves performing the feature measurements as described above, but we now replace the fitness function $F(\boldsymbol{x})$, which is a random variable, with a constant estimate $\hat{F}(\boldsymbol{x})$ of the expected fitness landscape like so:

$$\hat{F}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} F(\boldsymbol{x}). \tag{2}$$

We will test this method's effectiveness by empirically examining the relationship between the observed error in feature estimates and the number of samples N.

To test **RQ3**, we replace the $(1+1)$-style EA used in the feature calculations with a $(\mu + \lambda)$-style EA. We vary the value of μ and keep $\lambda = \mu$.

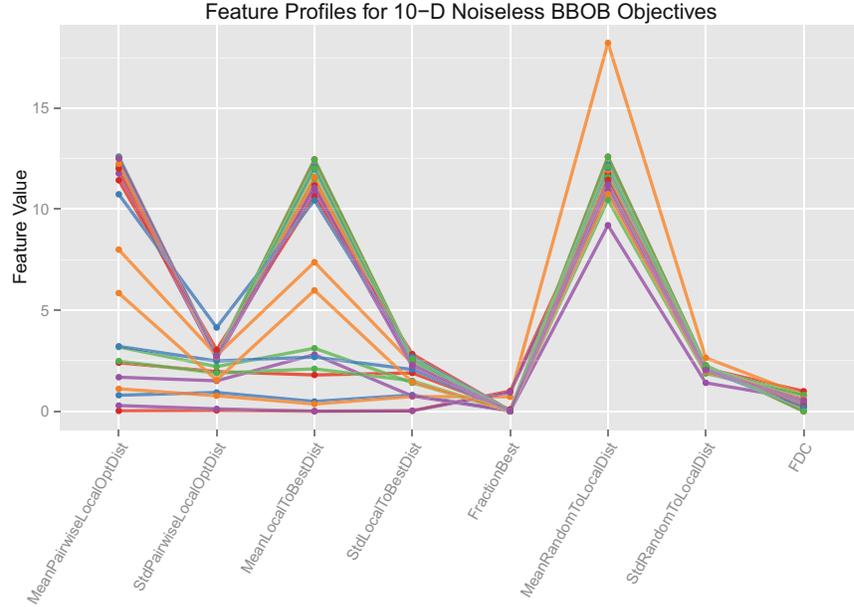Feature Profiles for 10−D Noiseless BBOB Objectives

Fig. 1. Mean feature profiles, averaged across 50 ten-dimensional instances of each of the 24 noiseless test functions in the BBOB suite.

3 Results

3.1 Sensitivity Analysis

The features we have implemented provide us with an eight-dimensional characteristic profile of each test function. The parallel plot in Fig. 1 visualizes these profiles for all 24 of the noiseless test functions. We consider the profile calculated from each noiseless landscape to be the 'true' feature values. The question is how our estimate of those feature values incurs error as noise increases (**RQ1**).

Figure 2 shows the value of each feature estimate, averaged over all 24 test functions, as we increase the value of p (see Eq. 1). While there is a great deal of variance in behavior across the 24 functions (not shown), in general we find that features 1–4 are extremely sensitive to noise: the estimate becomes inaccurate as soon as p reaches a value of about 10^{-3}. The remaining feature estimators (6–8) appear to be reasonably robust to small amounts of noise—but they suddenly become inaccurate when p reaches a threshold of about 0.25.

This answers **RQ1**: The features under study are highly sensitive to noise in the fitness landscape.

3.2 Explicit Averaging

We've shown that we can make the error in feature estimation explode by adding small amounts of artificial noise. Now we turn to the question of whether we can attenuate this error through explicit averaging of more than one fitness sample. We implemented explicit averaging for fitness evaluation during the hill-climber
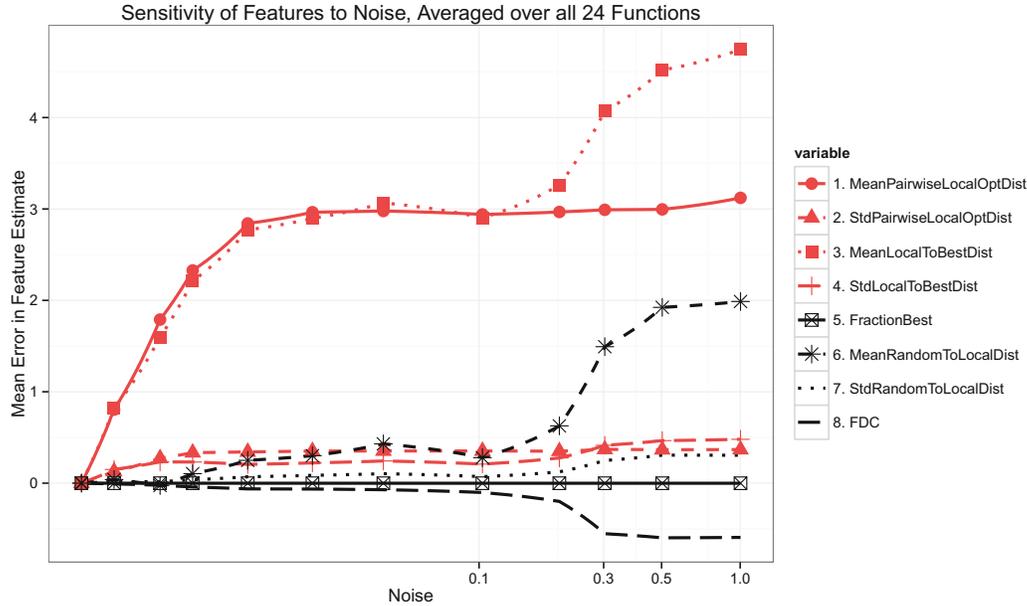
Fig. 2. Deviation between the estimated and true feature values as noise increases.

runs—that is, in this section we use the same $(1+1)$-style EA to compute feature estimates, but now every time an individual has its fitness evaluated, some $N \geq 1$ fitness samples are taken and averaged according to Eq. 2.

Let the magnitude of the noise be fixed at the small value of $p = 5 \cdot 10^{-4}$. Figure 3 shows the result of feature estimation averaged over all 24 instances while allowing the number of samples N to vary. We find that using explicit averaging of fitness has very little discernible impact on the accuracy of fitness measurements. Even at $N = 15$, a great deal of error remains.

Our answer to **RQ2** is thus negative: it seems that explicit averaging is not an effective means of correcting for noise.

3.3   Population-Based Search

It is well known that population-based search methods can perform a kind of 'implicit averaging' that makes their performance robust to noise. This is borne out in our experiments with the $(\mu + \lambda)$-EA, shown in Fig. 4. We see a sharp reduction in error when we increase μ from 1 to 2. As μ grows, however, we see stark, systematic deviations from the true feature values. This may be because, while the population-based EA is not significantly affected by small amounts of noise, it also has a tendency to converge to high-quality or *global* optima instead of the *local* optima that the features based on the $(1+1)$-EA are designed to seek out.

Our answer to **RQ3** is mixed, then: Replacing the hill-climbers in these features with a population-based algorithm does overcome noise, but it changes

the kind of information that the features gather from the landscape. Whether this information is useful for prediction or not is a question that is beyond the scope of this study.
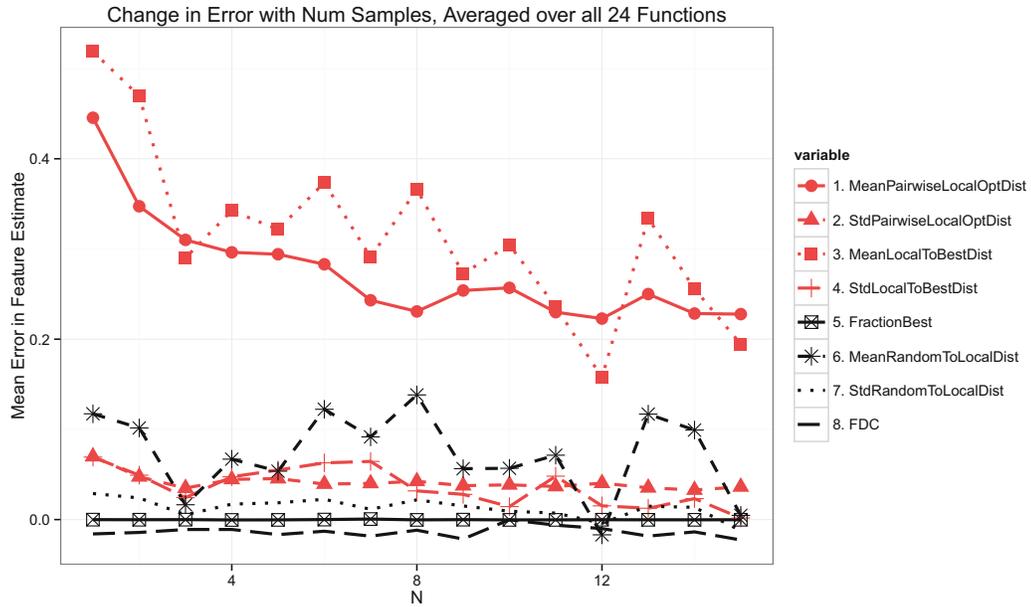

Fig. 3. Deviation between the estimated and true feature values as the number of explicit fitness samples increases.
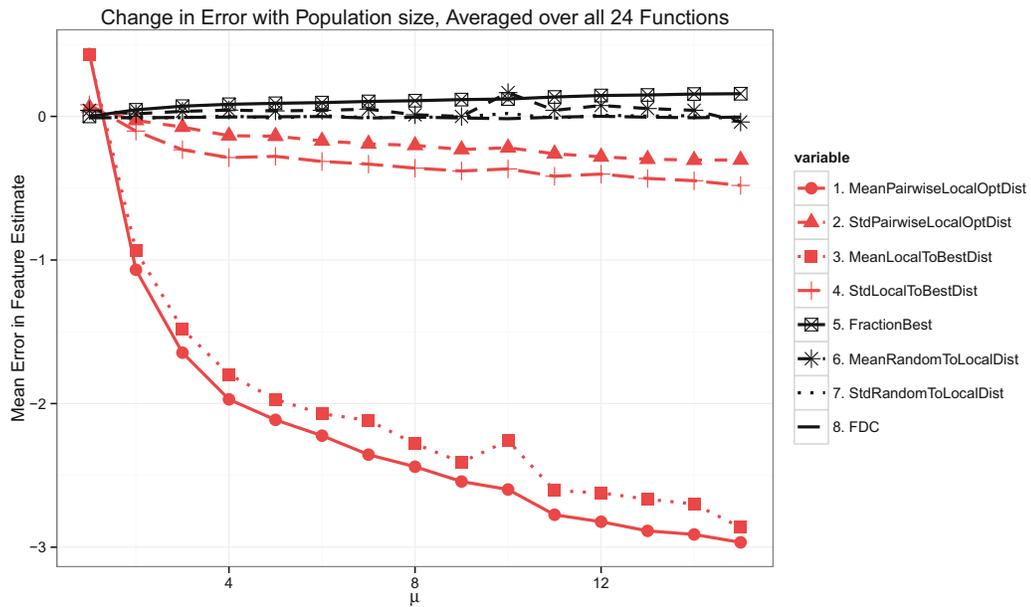

Fig. 4. Deviation between the estimated and true feature values as the population size increases.

4   Conclusion

Noise poses a particularly difficult challenge to the solution of computationally expensive problems. We find that the features under study, which are based on identifying local optima with a greedy search method, are sensitive to noise, yielding a positive answer to **RQ1**. Even very, very small quantities of noise are sufficient to entirely frustrate efforts to accurately measure features of a fitness landscape with these methods. Furthermore, we found that explicit averaging of many fitness samples is not sufficient to substantially attenuate the error caused by noise. A large number of fitness samples may be necessary to fully counteract even the impact that a very minuscule quantity of noise has on feature measurements—answering **RQ2** in the negative. Consequently, the features we have studied here, while they are initially appealing for computationally intensive applications because of their low cost, become computationally infeasible in the presence of noise.

We have shown that modifying these features to use a population-based algorithm in place of the hill-climbers is a promising approach, allowing us to overcome the issue of noise (**RQ3**). Because these algorithms are less greedy than a hill-climber, however, they gather different information about the landscape, and are less effective at collecting a representative sample of diverse local optima.

Our findings suggest that landscape analysis researchers should look toward the design of features that use population-based algorithms to gather information about the landscape. Future work might, for instance, explore replacing hill-climbers with state-of-the-art multimodal optimization methods. These may be able to overcome the noise problem while still gathering a representative sample of local optima [7]. Such an approach may be able to maintain some of the computational efficiency of the hill-climbing approach while also attaining some robustness to moderate amounts of noise.

References

1. Abell, T., Malitsky, Y., Tierney, K.: Features for exploiting black-box optimization problem structure. In: Nicosia, G., Pardalos, P. (eds.) LION 7. LNCS, vol. 7997, pp. 30–36. Springer, Heidelberg (2013)
2. Bassett, J.K.: Methods for improving the design and performance of evolutionary algorithms. Ph.D. thesis, George Mason University, Fairfax, VA (2012)
3. Beyer, H.-G.: Evolutionary algorithms in noisy environments: theoretical issues and guidelines for practice. Comput. Methods Appl. Mech. Eng. **186**(2), 239–267 (2000)
4. Bischl, B., Mersmann, O., Trautmann, H., Preuß, M.: Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In: Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation, pp. 313–320. ACM (2012)

5. Blum, C., Puchinger, J., Raidl, G.R., Roli, A.: Hybrid metaheuristics in combinatorial optimization: a survey. Appl. Soft Comput. **11**(6), 4135–4151 (2011)
6. Carlson, K.D., Nageswaran, J.M., Dutt, N., Krichmar, J.L.: An efficient automated parameter tuning framework for spiking neural networks. Front. Neurosci. **8**(10), 168 (2014)
7. Das, S., Maity, S., Bo-Yang, Q., Suganthan, P.N.: Real-parameter evolutionary multimodal optimization–a survey of the state-of-the-art. Swarm Evol. Comput. **1**(2), 71–88 (2011)
8. Deb, K., Goldberg, D.E.: Sufficient conditions for deceptive and easy binary functions. Ann. Math. Artif. Intell. **10**(4), 385–408 (1994)
9. Forrest, S., Mitchell, M.: What makes a problem hard for a genetic algorithm? some anomalous results and their explanation. Mach. Learn. **13**(2–3), 285–319 (1993)
10. Jin, Y., Branke, J.: Evolutionary optimization in uncertain environments-a survey. IEEE Trans. Evol. Comput. **9**(3), 303–317 (2005)
11. Jones, T., Forrest, S.: Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In: Sixth International Conference on Genetic Algorithms (ICGA 1995), vol. 95, pp. 184–192 (1995)
12. Kallel, L., Schoenauer, M.: Alternative random initialization in genetic algorithms. In: Seventh International Conference on Genetic Algorithms (ICGA 1997), pp. 268–275 (1997)
13. Manderick, B., de Weger, M., Spiessens, P.: The genetic algorithm and the structure of the fitness landscape. In: Proceedings of the 4th International Conference on Genetic Algorithms, pp. 143–150. Morgan Kaufmann, San Mateo, CA (1991)
14. Naudts, B., Kallel, L.: A comparison of predictive measures of problem difficulty in evolutionary algorithms. IEEE Trans. Evol. Comput. **4**(1), 1–15 (2000)
15. Smith-Miles, K.A.: Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Comput. Surv. (CSUR) **41**(1), 6 (2008)
16. Tayarani, M.-H., Yao, X., Hao, X.: Meta-heuristic algorithms in car engine design: a literature survey. IEEE Trans. Evol. Comput. **19**(5), 609–629 (2015)
17. Van Geit, W., De Schutter, D., Achard, P.: Automated neuron model optimization techniques: a review. Biol. Cybern. **99**(4–5), 241–251 (2008)
18. Vassilev, V.K., Fogarty, T.C., Miller, J.F.: Information characteristics and the structure of landscapes. Evol. Comput. **8**(1), 31–60 (2000)
19. Watson, J.P.: An introduction to fitness landscape analysis and cost models for local search. In: Gendreau, M., Potvin, J.Y. (eds.) Handbook of Metaheuristics. International Series in Operations Research & Management Science, vol. 146, pp. 599–623. Springer, Heidelberg (2010)