TECHNOLOGICAL AND ECONOMIC DEVELOPMENT OF ECONOMY
ISSN 2029-4913 / eISSN 2029-4921

2015 Volume 21(6): 899–916
doi: 10.3846/20294913.2015.1107654

SIMULATION-BASED FITNESS LANDSCAPE ANALYSIS AND OPTIMISATION OF COMPLEX PROBLEMS

Galina MERKURYEVA, Vitaly BOLSHAKOV

Department of Modelling and Simulation, Riga Technical University, 1 Kalku str., LV-1658 Riga, Latvia

Received 09 February 2015; accepted 09 October 2015

Abstract. Widespread hard optimisation problems in economics and logistics are characterised by large dimensions, uncertainty and nonlinearity and require more powerful methods of stochastic optimisation that traditional ones. Simulation optimisation is a powerful tool for solving these problems. Moreover, fitness landscape analysis techniques provide an efficient approach to better selection of a suitable optimisation algorithm. The concept and techniques of fitness landscape analysis are described. A formalised scheme for simulation optimisation enhanced with fitness landscape analysis is given. Benchmark fitness landscape analysis is performed to find relations between efficiency of an optimisation algorithm and structural features of a fitness landscape. Case study in simulation optimisation of vehicle routing and scheduling is described. Various optimisation scenarios with application of the fitness landscape analysis are discussed and investigated.

Keywords: optimisation, simulation, fitness landscape analysis, routing, scheduling, time windows.

JEL Classification: C60, C61, C63, L92.

Introduction

Nowadays solution of many problems in economics and logistics encounters challenges of high complexity and difficulties of the considered systems (Sakalauskas, Zavadskas 2009; Masri 2014). A factor that strongly influences the hardness of an optimisation problem is computational complexity of the problem. Different complexity (e.g., P, NP) classes characterise how computational time or space is dependent on the size of the problem input data (Garey, Johnson 1979; Sipser 2006). The P (polynomial time) complexity class contains decision problems if at least one optimisation algorithm exists that can solve the problem by a Deterministic Turing machine using a polynomial amount of computation. For the NP (nondeterministic polynomial time) problems, such an algorithm is unknown, but the

Corresponding author Galina Merkuryeva
E-mail: *galina.merkurjeva@rtu.lv*

problem could be solved in polynomial time by a nondeterministic Turing machine. More complexity classes distinguish optimisation problems solvable in the logarithmic or exponential time or with exponential memory resources. So-called NP-hard complexity class (Sipser 2006) contains problems that are computationally hard and are at least as hard as the NP-complete problems that are the hardest problems in NP.

NP-hard optimisation problems occur in economics in a variety of cases (Garey, Johnson 1979), such as routing in logistics, assignment tasks in manufacturing planning, scheduling (Eliiyi *et al.* 2009) in production and logistics, and supply chain planning (Napalkova, Merkuryeva 2012). Often, traditional optimisation methods (linear programming, integer programming, stochastic optimisation, etc.) could not be applied to solve hard optimisation problems. These methods may lead to ineffective solutions due to a high number of optimised parameters, stochastic nature of the optimised system and a large search space. A number of metaheuristic optimisation techniques are applied for the optimisation of these tasks. To choose an appropriate technique, fitness landscape analysis of the optimisation problem can be performed. Moreover, simulation of the system allows evaluating the system performance without analytical calculations. At present, simulation optimisation technology is a necessary tool for optimisation of complex systems, where evaluation of solutions can be complicated. Simulation-based fitness landscape analysis provides an efficient approach to analysis of suitability of an optimisation algorithm.

Nowadays, fitness landscape analysis methods are used for the determination of the problem hardness for metaheuristic algorithms (Stadler 2002; Pitzer, Affenzeller 2012; Pitzer *et al.* 2012a). However, there is a lack of information in literature on application of simulation in fitness landscape analysis within simulation optimisation of complex systems. Simulation-based fitness landscape analysis will allow better selection of an optimisation algorithm or construction of the most appropriate algorithm and its adjustment. The paper will discuss techniques for fitness landscape analysis and simulation optimisation of complex decision problems.

1. Simulation-based optimisation for complex systems

Modern optimisation problems in logistics are characterised by large dimensions, uncertainty and nonlinearity. Thus they require more powerful methods in stochastic optimisation than traditional ones, such as non-linear-programming methods or classical algorithms in stochastic dynamic programming. As mentioned above the main factors that strongly influence the hardness of the optimisation problem are computational complexity of the problem, system dynamics and stochastic nature of decision variables and parameters as well as difficulties in obtaining an analytical form of the objective function. To find solutions to such complex, large-scale, stochastic optimisation problems simulation-based optimisation approach is applied. In this case, optimisation methods are able to evaluate solutions by using simulation models instead of analytical expressions. Here, simulation technique is used to capture all the complexities and dynamics of the modelled system or optimised problem, whereas optimisation techniques are aimed at finding optimal or near-optimal solutions without explicitly evaluating each solution with a simulation model (Merkuryeva *et al.* 2011).

The review of optimisation methods traditionally used in simulation optimisation is given in (Visipkov *et al.* 1994). Gradient-based Search Methods estimate the simulation response function gradient to assess the shape of the objective function and employ deterministic mathematical programming techniques. Stochastic Optimisation methods find a local optimum for an objective function with stochastic values which are not known analytically. Statistical simulation optimisation methods use some additional information on the problem and structure of the simulation model. The Response Surface Methodology is based on approximation of a simulation response surface function by regression metamodels that fit the output variable of a simulation model in a small region of input factors. The regression metamodel's being an algebraic model of a simulation model has several advantages, such as explicit form, deterministic response and computational efficiency. Reproducibility of metamodels and statistical significance of solutions make RSM applicable for solving simulation optimisation tasks (Merkuryeva 2005). Most of these methods require an ability to estimate a gradient of the simulation response function, are designed for continuous optimisation problems and are hardly applicable for combinatorial problems.

The literature provides a number of numerical optimisation methods that need only the numerical value of the objective function for any solution candidate. These methods form a natural choice in solving complex stochastic optimisation problems, where the closed form of the objective function is frequently unknown (Gosavi 2003). Numerical methods also include metaheuristic optimisation methods, which facilitate finding good solutions for large and complex problems in a reasonable time with application of different heuristic and stochastic methods.

Although metaheuristic methods don't guarantee that the optimal solution to the problem will be found, there is a high interest in such methods in the applied optimisation of real-life problems (Glover, Kochenberger 2003). Let's note that heuristics provide rules for search algorithms to explore good solutions and avoid poorer solutions. Particularly, basic heuristics include random search and local search algorithms, and they are usually unable to locate global optimal solutions in case of a search space with multiple local optima. Metaheuristics (Dreo *et al.* 2006) coordinate an interaction between basic heuristics and higher level strategies to create an optimisation process that is capable of escaping from a local optimum and moving on to find other hopefully better local optima. These search methods are acknowledged as most applicable ones for solving simulation optimisation problems.

The main classes of metaheuristic optimisation methods are Genetic Algorithm (GA) (Goldberg 1989), Evolution Strategy (ES) (Schwefel 1995), Simulated Annealing (Kirkpatrick *et al.* 1983) and Tabu Search (Glover 1989). The application of the metaheuristic and other numerical methods becomes more important for especially NP-hard combinatorial optimisation problems (Dreo *et al.* 2006).

To make a selection and adjustment of an optimisation method more reasonable, a fitness landscape analysis offers specific techniques for the investigation of the problem search space.

2. Fitness landscape analysis review

Fitness landscape analysis provides methods and techniques for a mathematical analysis of a search space of optimisation problems. It can be applied as a support tool to enhance optimisation of complex problems, and it is widely considered in literature (Weinberger 1990; Jones, Forrest 1997; Stadler 2002). A fitness landscape is interpreted as a combination of a fitness function of an optimisation problem and relationships or a distance metric between solutions in a search space (Reeves, Rowe 2002). It is proposed that the structure of the fitness landscape affects the way in which a search space is examined by a metaheuristic optimisation algorithm. Fitness landscape analysis would allow getting more information on the problem's properties dependent on a specific optimisation method, which will guide the optimisation process (Reeves, Rowe 2002). With the landscape analysis it is possible to get measures of the problem's difficulty and find recommended configuration of an optimisation algorithm (Beham *et al.* 2013; Pitzer *et al.* 2013). Moreover, searching for better specific algorithms and their configurations for the problem subclasses provides useful knowledge on the problem solution scenarios (Pitzer, Affenzeller 2012).

Formally, a fitness landscape can be defined (Jones 1995) as follows. Let the representation space R denote a set of representations and a search operator is a function ϕ: $M(R) \times M(R) \to [0, 1]$, where $M(R)$ is a multiset of representations. For representations $v, w \in M(R)$, $\phi(v, w)$ defines a probability p that v will be modified to w by application of operator ϕ. Fitness landscape L is defined by a 5-tuple:

$$L = (R, \phi, f, F, >_F), \tag{1}$$

where f is a fitness function; F is a fitness space with partial order $>_F$. The landscape can also be represented as a directed labelled graph $G_L = (V, E)$, where vertices are $V \subseteq M(R)$, and edges are $E \subseteq V \times V$. In this representation, a vertex $v \in V$ is labelled as $f(v)$, and an edge (v, w) is labelled $\phi(v, w)$. Similar to nature landscapes, hill ridges, valleys and other structures can be identified on the fitness landscapes and formalised as peak (or optimum), local-optimum, global-optimum, plateau and basin of attraction.

There are several characteristics associated with the landscape optima that define the structure of fitness landscapes and affect mostly the hardness of an optimisation problem. They are: the modality, which defines a number and density of optima in a search space (Reeves, Rowe 2002); ruggedness that characterises the impact of all landscape structures on the hardness of the search (Merz, Freisleben 2000); and neutrality, which defines a number of plateaus (Reidys, Stadler 2001).

A number of different techniques and metrics have been developed for analysis of fitness landscapes by evaluating their structural characteristics (Hordijk 1996; Jones, Forrest 1997; Vassilev *et al.* 2000; Smith *et al.* 2002; Collard *et al.* 2004; He *et al.* 2004; Vanneschi *et al.* 2006; Czech 2008). They do not require information about all problem solution candidates, but analyse only a part of a fitness landscape data and apply different strategies for data collection. These strategies are based on simple moves, which generate a trajectory through the landscape. For example, in the random walk, a solution candidate is randomly and repeatedly modified. In the adaptive walk, a certain number of mutations are performed to generate a set of neighbours, and then the best one is selected (Kauffman 1989). The up-down walk is similar to the adaptive one, but the walk is in reverse direction when a local optimum is reached (Vassilev *et al.* 2000). Finally, neutral walks explore "flat" areas (Reidys, Stadler 2001). It is supposed that reviewed fitness landscapes are statistically isotropic. A technique for isotropy measurement is proposed in (Pitzer *et al.* 2011).

In statistical analysis techniques (Weinberger 1990), the autocorrelation function is used to measure the ruggedness of the landscape. In case of a low autocorrelation between two sets of fitness points separated with some solutions, these points have dissimilar values, and the landscape is more rugged. Fitness values f_t obtained by the random walk on the landscape form a sequence or time series $\{f_t\}_{t=1}^{N}$ of length N. Autocorrelation function $\rho(\Gamma)$ between sets of fitness points separated by a distance Γ is calculated by:

$$\rho(\Gamma) \approx \frac{E(f_t f_{t+s}) - E(f_t)E(f_{t+s})}{V(f_t)}, \tag{2}$$

where $E(f_t)$ is the expectation and $V(f_t)$ is the variance of a sequence $\{f_t\}_{t=1}^{N}$. For smooth landscapes the autocorrelation of fitness points in a random walk is close to 1 and tends to zero for rugged landscapes (Reeves, Rowe 2002). Another correlation metric used in practice is correlation length τ that defines a distance beyond which two sets of fitness points become uncorrelated:

$$\tau = \frac{1}{\ln(\rho(1))}. \tag{3}$$

A longer correlation length indicates a smoother landscape while a shorter one would indicate a more rugged landscape. More robust estimation of the correlation length is proposed by Hordijk (1996). It is extended by the definition, that correlation is significant while it exceeds two standard-error-bounds $\left(-2/\sqrt{N}; +2/\sqrt{N}\right)$.

In information analysis techniques, the concept of entropy proposed in classical information theory is used as a basic concept to quantify the ruggedness of a landscape. The fitness landscape is interpreted as an ensemble of objects, which are characterised by their form, size and distribution. These objects consist of a point on the fitness landscape and its nearest neighbours. All information measures are calculated with notice to a calculation accuracy that is defined by parameter ε that defines threshold slopes in the fitness path. All slopes that have fitness difference between the neighbour solutions less than ε are assumed to be flat (Vassilev *et al.* 2000).

Four information measures are proposed in literature. Information content $H(\varepsilon)$ and partial information content $M(\varepsilon)$ are two measures of the entropy or amount of fitness change encountered during the walk in the obtained landscape path. They indicate the ruggedness and the modality of the landscape path correspondingly. Particularly in case of high information content, the landscape has a large variety of structures and is more rugged. The information stability ε^* characterises a magnitude optimum in the obtained landscape fitness path. The density-basin information $h(\varepsilon)$ analyses the variety of flat and smooth sections on the landscape.

To obtain these information measures of the landscape L, the landscape walk is performed, and the sequence of $\{f_t\}_{t=1}^{N}$ collected fitness values is transformed into a string of ensembles $S(\varepsilon)$, with elements $s_t \in \{\bar{1},0,1\}$. Information content $H(\varepsilon)$ is defined by:

$$H(\varepsilon) = -\sum_{p \neq q} P_{[pq]} \log_6 P_{[pq]}, \tag{4}$$

where the parameter ε controls the sensitivity for measuring the entropy $H(\varepsilon)$ and is a real number from the interval [0, £], where £ is the maximum fitness difference of the sequence $\{f_t\}_{t=1}^{N}$. Probabilities $P_{[pq]}$ represent frequencies of possible sub-blocks *pq* of elements from string $S(\varepsilon)$. Partial information content $M(\varepsilon)$ is determined by:

$$M(\varepsilon) = \frac{\Phi_s(1,0,0)}{n}, \tag{5}$$

where $\Phi_s(1,0,0)$ counts the slopes of the optima, that are represented by string $S(\varepsilon)$ of length equal to N. For further explanations we refer to (Vassilev *et al.* 2000). Finally, the density-basin information $h(\varepsilon)$ is determined by:

$$h(\varepsilon) = -\sum_{p \in \{1,0,\bar{1}\}} P_{[pp]} \log_3 P_{[pp]}, \tag{6}$$

where probabilities $P_{[pp]}$ represent frequencies of sub-blocks *pp* from the string $S(\varepsilon)$ (Vassilev *et al.* 2000). Information stability ε* is the lowest ε value, when a fitness path has no structures at all.

3. Simulation-based fitness landscape analysis and optimisation

To extend the concept of a fitness landscape for its application in simulation optimisation, fitness landscape L' is introduced where fitness of solutions is evaluated using a simulation model instead of an analytical expression:

$$L' = (R, \phi, S). \tag{7}$$

Here, R is a representation space, ϕ is a search operator and S denotes a simulation model with one output variable:

$$S = f : \vec{x} \times \xi \mapsto \hat{y}, \tag{8}$$

where $\vec{x} \in R$ is a vector of simulation model input variables, which represent solution candidates in the representation space; f is an objective function, ξ is a random component of the model, and $\hat{y} \in \mathbb{R}$ is the mathematical expectation of the simulation model output.

To apply fitness landscape analysis in simulation optimisation, the following three-level formalised scheme (see Fig. 1) is introduced (Bolshakov 2013). At the benchmarking level, information on landscape measures and performance of optimisation algorithms on benchmark landscapes is collected. At the landscape analysis level, the landscape analysis procedure is performed. The trajectory on the landscape is generated with different walking strategies, the time series of fitness values are obtained and statistical and information measures for fitness landscape analysis are calculated. The obtained data are used to select and adjust an appropriate optimisation algorithm. At the optimisation level, the selected algorithm is used to solve the problem using a simulation-based metaheuristic optimisation approach.

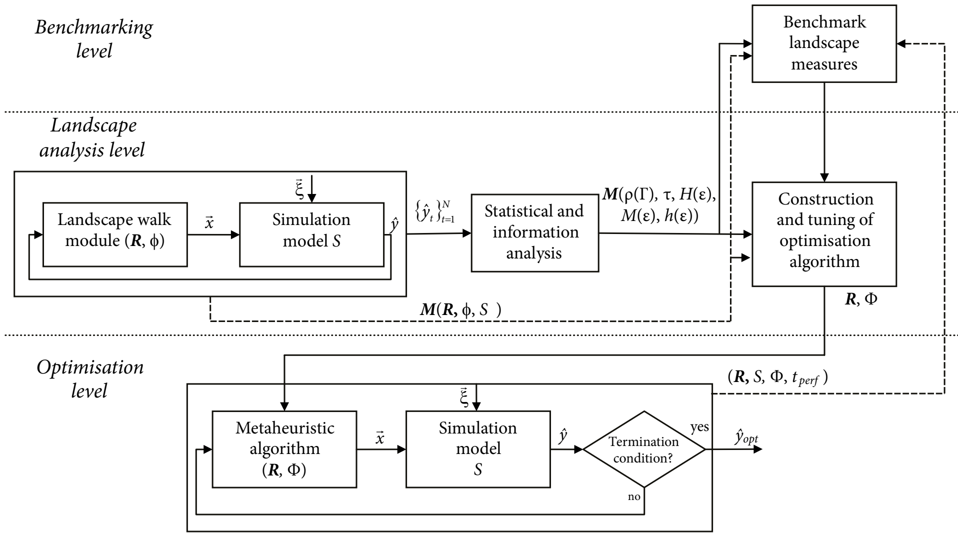

Fig. 1. Simulation optimisation with fitness landscape analysis

Let us define main components of the proposed scheme. *Simulation model S* is used to evaluate the performance of a system to be optimised. It produces the simulation output from several model runs or replications and is defined by:

$$S = f : \vec{x} \times \vec{\xi} \mapsto \hat{y}, \tag{9}$$

where f is an objective function, with which an optimal value is searched for in the optimisation process; $\vec{x} = (x_1, ..., x_i, ..., x_k)$ is a vector of k input variables, $\vec{x} \in R$; $\vec{\xi} = (\xi_1, \xi_2, ..., \xi_d)$ is a disturbance vector of d environmental variables; $\hat{y} = E\left[y\right]$ is a mathematical expectation of simulation output data, where $y \in \mathbb{R}$ is simulation output in each replication. For each input variable it is true that $x_i \in \mathbb{R}$, $0 \le i < k$. Here, the simulation model is interpreted as a black box that defines input-output relationships of the model without identifying its states. *Landscape walk module LW* is formalised as follows:

$$\vec{x}_{t+1} = LW(\vec{x}_t, \phi, \hat{y}), \tag{10}$$

where t is a number of iterations completed in a walk, and $\vec{x}_{t+1}$, $\vec{x}_t$ define vectors of input variables to simulation model S in the current and previous iterations. Here, $\hat{y}$ is used as module input in up-down and neutral landscape walks for determining walking direction by using an operator ϕ, and a vector $\vec{x}$ is module output. As a result of process integration of modules LW and S, time series $\left\{\hat{y}_t\right\}_{t=1}^{N}$ with a number N of evaluations in the trajectory are generated that define sequences of fitness values denoted above by $\left\{f_t\right\}_{t=1}^{N}$. The module of *statistical and information analysis* (*S&IA*) based on the obtained time series $\left\{f_t\right\}_{t=1}^{N}$ calculates a set of the landscape measures $\rho(\Gamma)$, τ, $H(\varepsilon)$, $M(\varepsilon)$, $h(\varepsilon)$ for different values of autocorrelation distance Γ and sensitivity ε, i.e.:

$$\left(\left\{f_t\right\}_{t=1}^{N}\right) \times \Gamma \times \varepsilon \mapsto \left(\rho(\Gamma), \tau, H(\varepsilon), M(\varepsilon), h(\varepsilon)\right). \tag{11}$$

The module for *construction and tuning of an optimisation algorithm* allows selecting an appropriate metaheuristic optimisation algorithm, its configuration, and adjusting its parameters for a specific problem. Data received at benchmarking and landscape analysis levels are used as inputs, and model output defines a representation space R and a set Φ of search operators that form an optimisation algorithm. It is based on the rules and recommendations applicable for known values of the landscape measures.

At the optimisation level, the *metaheuristic algorithm* (*MA* module) produces the best solutions in the automatic search process using a simulation model to evaluate solution candidates. This module can be formalised as follows:

$$MA : (R, \Phi, \hat{y}, M(\langle \vec{x}, \hat{y} \rangle)_t) \mapsto \vec{x}_{t+1}, \tag{12}$$

where $M(\langle \vec{x}, \hat{y} \rangle)_t$ denotes a memory of previous solution candidates obtained after the t-th evaluation. When the optimisation cycle defined by a *termination condition* is completed, the best found solution $\hat{y}_{opt} = \langle \vec{x}, \hat{y} \rangle$ is selected. The performance measures of the optimisation model can be added to the dataset of benchmark landscape measures including the problem landscape L' and time performance t_{perf} of the optimisation algorithm.

4. Experimental analysis for benchmark fitness landscapes

The following fitness function0s used for benchmarking of genetic algorithms (De Jong 1975; Rastrigin 1974; Ackley 1987) were experimentally analysed, i.e., Sphere, Rosenbrock, Rastrigin and Ackley functions. All these functions can be defined in the same search domain with a similar number of variables and can easily be graphically interpreted for two variables. For example, so called Rastrigin function has more rugged landscape and in case of n variables is defined by:

$$f_{Rastrigin}(\langle x_1, ..., x_n \rangle) = 10n + \sum_{i=1}^{n} \left(x_i^2 - 10\cos(2\pi x_i) \right). \tag{13}$$

In case of the function minimisation, it has one global optimum in point of $x_i = 0, i = 1, ..., n$ with the value $f_{Rastrigin} = 0$. But this function is highly multimodal having the local optima produced by its cosine component (see Fig. 2).

For all four functions, the number of variables is taken equal to $n = 2$, and a search domain is defined by $-5 \le x_i \le 5$, $i = 1, 2$. Two types of solution representation are used, i.e. real-value encoding and binary representation. In the first case variables x_1 and x_2 are coded as real numbers with a resolution factor of 0.01. Binary coded chromosomes have length of 20, where first 10 bits code x_1, but others code x_2. For real-value encoding, a mutation operator is applied that changes each variable in a chromosome by +0.01 or −0.01 with a probability equal to 1/3. For binary representation, a bitflip operator changes a value of a randomly selected bit to an opposite one. So, eight different fitness landscapes are analysed, i.e. four different benchmark functions and two types of search space for each function.

Results of multiple global and local landscape experiments performed to estimate their structural measures using developed software are given in (Merkuryeva, Bolshakovs 2011). Particularly, the effects of the length generated by a random walk on statistical and information measures of benchmark landscapes and a random noise in fitness were analysed. To define the effect of a random noise in a fitness function, these experiments were performed with a modified fitness function f^* such as:

$$f^* = f + \xi, \tag{14}$$

where f is the true fitness function, and ξ is a term that represents noise effects that is treated as a statistical error and assumed to be normally distributed with mean zero and variance σ^2. The results of experiments show that both statistical and information measures are quite sensitive to noise. Furthermore, with increase of variance the autocorrelation gets lower for shorter lags and higher for longer lags. Additionally, a random noise increases the entropy of the fitness landscape structures.

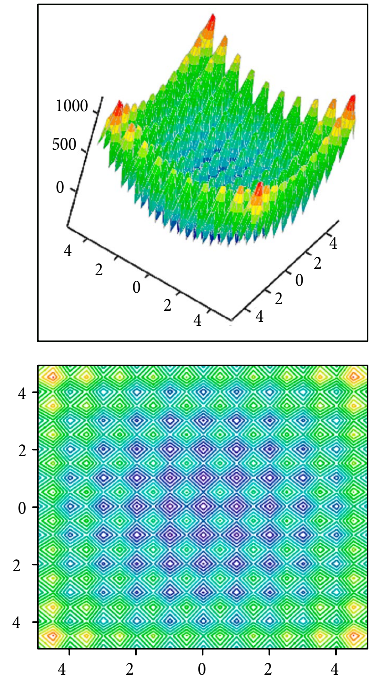

Fig. 2. Rastrigin function with 2 variables and its contour plot

Nowadays, a toolkit for comprehensive and automatic fitness landscape analysis execution (Pitzer *et al.* 2011) implemented within the open-source optimisation software HeuristicLab allows plugging typical fitness landscape analysis techniques into existing algorithms and provides information about fitness landscape characteristics during execution. Moreover, in (Pitzer *et al.* 2012b) fitness landscape analysis is performed to determine in what way problem instances of the same class (e.g., vehicle routing problems) are similar or different. All values of landscape measures obtained by different operators were joined in a vector of problem instances characteristics and analysed to perform cluster analysis of the problem instances.

5. Simulation optimisation of vehicle schedules

Vehicle Scheduling Problems (VSP) present a class of optimisation problems aimed at assigning a set of scheduled trips to a set of vehicles, that each trip is associated with one vehicle, and a cost function for all trips is minimised (Eliiyi *et al.* 2008; Nagamochi, Ohnishi 2008). The problem is often modified with additional constraints, like time windows, different vehicle capacity, etc (Merkuryeva, Bolshakov 2015). A number of methods to solve VSP problems are proposed in literature, e.g. integer programming, combinatorial methods, heuristics. In practice, the VSP also can be complicated by stochastic processes existing in the problem, e.g. when the duration of a trip is a random variable. In this case, evaluation of potential solutions can be made through simulation, and simulation optimisation could be used to solve such problems.

5.1. Problem statement

A vehicle schedule defines a schedule of deliveries of various types of goods from a distribution centre (DC) to a network of stores (Merkuryeva, Bolshakov 2014). Distribution routes or trips for vehicles are fixed. For each trip, a sequence of stores (points), average time intervals for vehicles moving between these points, loading and unloading and types of goods to be carried are defined. Goods are delivered to stores in the predefined time windows. For each store, an average demand of goods of each type is defined. Vehicle capacities are limited and known. Vehicles are assigned to routes and schedules for routes are generated that minimise the total costs of a schedule. The vehicle idle time is defined as a sum of time periods, when a vehicle is waiting for the next trip.

Decision variables are introduced to assign vehicles v_i to routes and define a start time t_i, for each route, where i is a route number, v_i is a vehicle assigned to trip i and t_i is start time of the trip i. The problem constraints are vehicle capacity constraints, delivery time windows and gate capacity constraints. Express analysis shows that the problem could have many solutions that are not feasible within defined constraints. This makes a solution search process non-efficient in terms of computational time. To increase optimisation efficiency, all constraints are converted into soft constraints (Merkuryeva, Bolshakovs 2010), and the objective function f is specified as follows:

$$f = \sum_{i=1}^{N} T_{idle}^{i} + k_1 T_c + k_2 T_m + k_3 T_0 + k_4 N_{ol} + k_5 N_{ot} \rightarrow \min, \tag{15}$$

where T_{idle}^{i} is the total idle time for vehicle i; N is a number of vehicles; T_c defines the total duration of overlapping trips for one vehicle; T_m defines the total time of window mismatches; T_o and N_{ol} determine the total time and a number of vehicles that exceeded the total working time; and N_{ot} is a number of vehicles overloaded. All indexes for unsatisfied constraints are multiplied with penalty coefficients k_i.

5.2. Fitness evaluation and analysis through simulation

To estimate fitness of potential vehicle schedule solutions, a discrete-event simulation model in AnyLogic simulation software is built (Merkuryeva, Bolshakovs 2010). The vehicle moving time between two points is defined by a normally distributed random variable. The efficiency of a schedule candidate is evaluated by the average total idle time of all vehicles, and during simulation constraints violations are monitored. To validate the model, a schedule from a case study was simulated.

The results of fitness landscape analysis for the problem with stochastic input data (stochastic times for moving, unloading, etc.) and deterministic data expressed by mean values are given in Table 1. In each series of simulation experiments, 5 replications with 100

Table 1. Information and statistical measures

Model input data	$H(0.1)$	$M(0.1)$	$h(0.1)$	ε^*	$\rho(1)$	$\rho(10)$	τ
Stochastic	0.66	0.20	0.49	0.40	0.84	0.21	7.24
Deterministic	0.62	0.17	0.37	0.35	0.89	0.32	8.75

solutions in the path were made. Sensitivity value ε for calculation of information measures is given as 0.1 of the difference between the smallest and largest fitness values in the path.

To perform a random walk on the fitness landscape, a mutation operator changes one randomly selected trip in the solution candidate by assigning a new randomly chosen vehicle, and start time is shifted to the later one by a certain constant value (Merkuryeva, Bolshakov 2012a). A sample random walk in one experiment is shown in Figure 3.

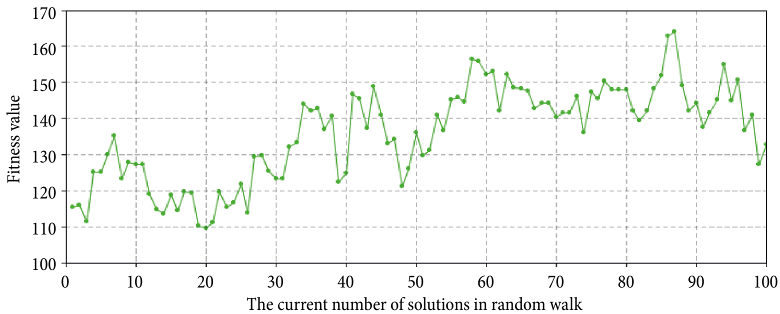

Fig. 3. Fitness values in a random walk on the landscape

The problem with stochastic data seems to be more complex for an optimisation algorithm as far as values of the autocorrelation function between adjacent solutions ρ(1) are lower, and information measures have higher entropy and higher modality.

Statistical measures indicate that both landscapes are relatively rugged. The autocorrelation ρ(10) between two sets of landscape points separated by 10 solutions is very low. Additionally, the information content $H(0.1)$ is relatively high. The partial information content $M(0.1)$ is low and as a result, the modality of landscapes should be low. Density-basin information $h(0.1)$ indicates that local peaks have high density. Results of additional experiments and a comparative analysis show that the defined problem landscape is less rugged than landscapes of the benchmark fitness functions whose solutions are coded in binary chromosomes. Thus, this problem could be solved using genetic algorithms (GA) as those problems based on the benchmark functions.

5.3. Optimisation scenarios

Fitness evaluation with simulation is time consuming. To perform a faster and more comprehensive analysis, the simulation model was re-implemented as a plug-in of HeuristicLab (Wagner 2009) by maintaining its logic. To enhance the quality of optimisation results, permutation encoding of solutions is introduced. A chromosome contains $m + n$ genes, where n is a number of vehicles and m is a number of trips. The genes that have values less or equal to m encode a trip number, and values greater than m encode delimiters or vehicle designators and define, that the next sequence of trips should be performed by the corresponding vehicle (Bolshakov *et al.* 2011).

A grid of landscape analysis experiments was created to compare values between different landscapes (Bolshakov *et al.* 2011). First, comparison of different mutation operators is performed, and, second, comparison between existing and proposed encodings is done. In particular, it was found that in a random walk, values of autocorrelation function are slightly lower for the replacement operator. In up-down walk the situation is opposite: replacement mutation has higher correlation than shift mutation, but three artificial problems are different to the others (see Figure 4, where dots are for replacement and crosses for shift mutators). Moreover, the value of the autocorrelation function in random and up-down walks is lower for the permutation encoding. It means that landscapes of this encoding should be more rugged.

A number of VSP optimisation algorithms implemented in the Heuristic framework, i.e. Evolution Strategy (ES), Simulated Annealing (SA) and GA, were tested. For integer encoding, both ES and SA algorithms are fast and highly successful, but the ES is able to find solutions with better quality (Fig. 5). A GA finds even better solutions, but requires larger numbers of evaluations. Permutation encoding is found to be more effective for optimisation of the VSP, as algorithms are able to find solution in less time. Moreover, almost all idle times are eliminated, and trip overlapping id also avoided. Even though the search space for this type of encoding is more complex and rugged, nevertheless, due to its smaller size the search of the globally optimal solution becomes more effective.

If the GA is selected, permutation encoding has to be chosen unless the problem contains more than 100 trips. In case of integer vector encoding, selection of an appropriate mutation operator is based on the statistical analysis: an operator with the highest autocorrelation value in the up-down walk should be selected.

Furthermore, an additional simulation optimisation scenario was investigated to define an optimal set of delivery routes and schedules of routed solutions. It is supposed, that an optimal distribution of routes and vehicles would minimise a number of required vehicles as well as decrease delivery distances and the total vehicle idle time. Correspondingly, two optimisation problems are introduced (Bolshakov 2013) and solved sequentially: 1) vehicle routing with time windows, and 2) vehicle route scheduling.

For vehicle routing, an island genetic algorithm with offspring selection (IOSGA) described in (Vonolfen *et al.* 2011) is applied. It presents a coarse-grained parallel GA where population is divided into several islands in which GA works independently. Periodically, after a certain number of generations best solutions migrate between islands. Offspring selection forces the algorithm to produce offspring solutions with better fitness than their parents (Affenzeller *et al.* 2009). For the considered problem instances, operators and parameters of the IOSGA were determined experimentally: a proportional selector; 5 islands; 200 individuals in population; ring migration each 20 generations with 15% rate: random individuals are replaced with the best ones from a neighbour island. The maximal selection pressure was set equal to 200 and the mutation rate equal to 5%. A GVR crossover (Pereira *et al.* 2002) was selected as it works with an unlimited number of vehicles, but provides best results in terms of keeping routes not overloaded.

It is worth mentioning that GA is not considered as the strongest optimisation method for the VRP (Gendreau *et al.* 2001; Baker, Ayechew 2003), and more often the Tabu Search (TS) algorithm with constraint relaxation (Cordeau *et al.* 2001) is recommended as more

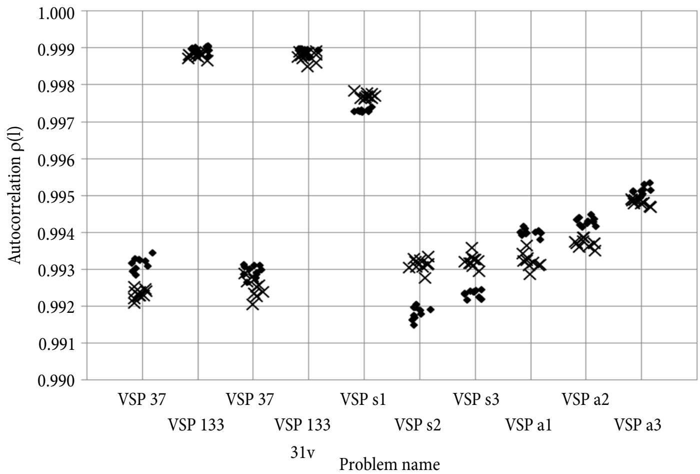

Fig. 4. Autocorrelation in up-down walks

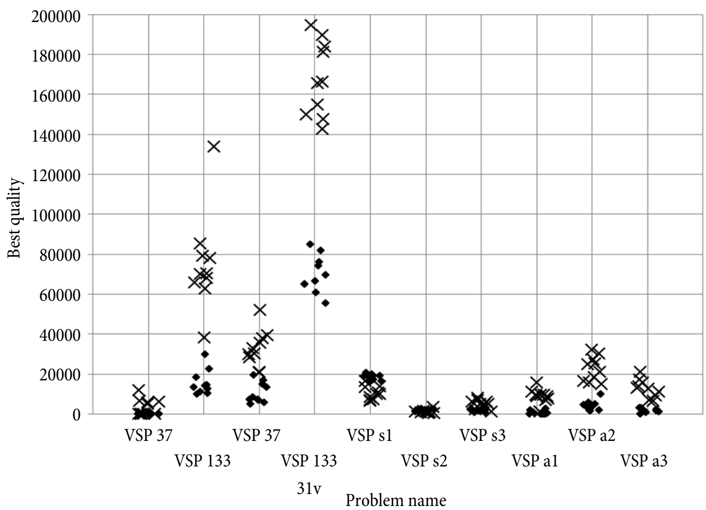

Fig. 5. Quality of best solutions found with ES

efficient approach. However, genetic algorithms show good performance for routing problems and are highly robust and adjustable. Also, for the considered example instance, the IOSGA showed better results than TS.

For route scheduling, permutation encoding (Merkuryeva, Bolshakov 2012b) is applied, different optimisation algorithms, such as ES, GA, Island GA with 5 islands and Offspring Selection GA (Affenzeller *et al.* 2009) were examined. Finally, the ES was selected for its ability to find the best results with fewer evaluations.

The results of a sample experiment for a day plan and specific demand data from 53 stores are presented in Figure 6. The best found routing solution defines 34 routes, which may be combined due to long time windows. Furthermore, the ES (20 + 100) algorithm

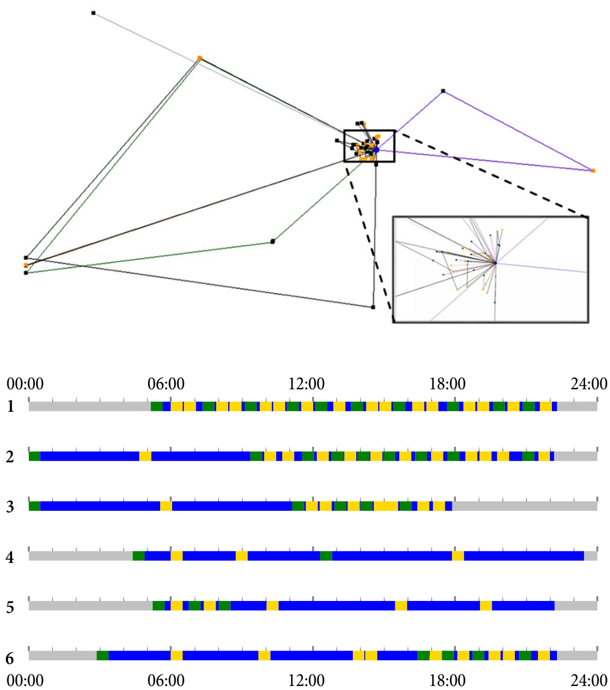

Fig. 6. Vehicle routing and scheduling solutions for a sample instance

was applied for scheduling of the routed solution. As a result, the optimal schedule was found with all constraints satisfied if at least 6 vehicles are available.

Vehicle routing optimisation complemented with scheduling of routed solution can be applied in the delivery tasks, where vehicle trips are relatively short in comparison with a planning horizon.

Conclusions

Complex optimisation problems in economics and logistics are characterised by large dimensions, nonlinearity and uncertainty, and require more powerful techniques in stochastic optimisation. The proposed formalised scheme for simulation-based fitness landscape analysis and optimisation allows extending the concept of a fitness landscape for its application in simulation optimisation. Evaluation of structural characteristics of fitness landscape provides data in order to select an appropriate optimisation algorithm as well as adjust its components and parameters. Experimental analysis of benchmark landscapes allows finding relations between structural features of fitness landscapes, their measures and behaviour of optimisation algorithms on these landscapes.

Two types of vehicle schedule optimisation solutions are developed in the case study. The first one is designed in way that has predefined vehicle routes and is simulation-based with stochastic factors. The second solution allows optimising both vehicle routes and schedules when routes are not predefined. Sequential vehicle routing and scheduling of routed solutions allows obtaining cost-effective solutions in large store network delivery planning by minimising a number of required vehicles and related costs. In both solutions, fitness landscape analysis is applied for investigating and adjustment of optimisation algorithms to be performed in each specific case.

Funding

Support for this work was provided by the Riga Technical University through the Scientific Research Project Competition for Young Researchers No. ZP-2014/21.

References

Ackley, D. H. 1987. An empirical study of bit vector function optimization, in L. Davis (Eds.). *Genetic algorithms and simulated annealing.* London: Pittman Publishers, 170–215.

Affenzeller, M.; Winkler, S.; Wagner, S.; Beham, A. 2009. *Genetic algorithms and genetic programming: modern concepts and practical applications.* Chapman and Hall/CRC. http://dx.doi.org/10.1201/9781420011326

Baker, B. M.; Ayechew, M. A. 2003. A genetic algorithm for the vehicle routing problem, *Computers and Operations Research* 30: 787–800. http://dx.doi.org/10.1016/S0305-0548(02)00051-5

Beham, A.; Pitzer, E.; Affenzeller, M. 2013. Fitness landscape based parameter estimation for robust taboo search, in R. Moreno-Díaz, F. Pichler, A. Quesada-Arencibia (Eds.). *EUROCAST 2013, Part I, LNCS* 8111: 292–299. Springer. http://dx.doi.org/10.1007/978-3-642-53856-8_37

Bolshakov, V. 2013. *Simulation-based fitness landscape analysis and optimisation of complex systems*: PhD thesis. Riga Technical University.

Bolshakov, V.; Pitzer, E.; Affenzeller, M. 2011. Fitness landscape analysis of simulation optimisation problems with heuristiclab, in *Proceedings of the UKSim 5th European Symposium on Computer Modeling and Simulation*, 16–18 November 2011, Madrid, Spain, 107–112. http://dx.doi.org/10.1109/EMS.2011.14

Collard, P.; Verel, S.; Clergue, M. 2004. Local search heuristics: fitness cloud versus fitness landscape, in *Proceedings of 16th European Conference on Artificial Intelligence,* 22–27 August 2004, Valencia, Spain, 973–974.

Cordeau, F.; Desaulniers, G.; Desrosiers, J.; Solomon, M. M.; Soumis, F. 2001. VRP with time windows, in P. M. Toth, D. Vigo (Eds.). *The vehicle routing problem.* Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 157–193.

Czech, Z. J. 2008. Statistical measures of a fitness landscape for the vehicle routing problem, in *IEEE International Symposium on Parallel and Distributed Processing*, 14–18 April 2008, Miami, Florida, USA, 1–8. http://dx.doi.org/10.1109/IPDPS.2008.4536369

De Jong, K. D. 1975. *An analysis of the behavior of a class of genetic adaptive systems*: PhD thesis. Department of Computer and Communication Sciences, University of Michigan.

Dreo, J.; Petrowski, A.; Siarry, P.; Taillard, E. 2006. *Metaheuristics for hard optimization. Methods and case studies*. Berlin Heidelberg: Springer-Verlag.

Eliiyi, D. T.; Korkmaz, A. G.; Cicek, A. E. 2009. Operational variable job scheduling with eligibility constraints: a radnomized constraint-graph-based approach, *Technological and Economic Developemnt of Economy* 15(2): 245–266. http://dx.doi.org/10.3846/1392-8619.2009.15.245-266

Eliiyi, D. T.; Ornek, A.; Karakutuk, S. S. 2008. A vehicle scheduling problem with fixed trips and time limitations, *International Journal of Production Economics* 117(1): 150–161. http://dx.doi.org/10.1016/j.ijpe.2008.10.005

Garey, M. R.; Johnson, D. S. 1979. *Computers and intractability. A guide to the theory of NP-completeness*. New York: W.H. Freeman and Company.

Gendreau, M.; Laporte, G.; Potvin, J. Y. 2001. Metaheuristics for the capacitated VRP, in P. Toth, D. Vigo (Eds.). *The vehicle routing problem*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 129–154.

Glover, F. 1989. Tabu search – Part I, *INFORMS Journal on Computing* 1(3): 190–206. http://dx.doi.org/10.1287/ijoc.1.3.190

Glover, F.; Kochenberger, G. A. 2003. *Handbook of metaheuristics*. International series in operations research & management science. Springer. http://dx.doi.org/10.1007/b101874

Goldberg, D. E. 1989. *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley Professional.

Gosavi, A. 2003. *Simulation-based optimization: parametric optimization techniques and reinforcement learning*. Kluwer Academic Publishers. http://dx.doi.org/10.1007/978-1-4757-3766-0

He, J.; Yao, X.; Zhang, Q. 2004. To understand one-dimensional continuous fitness landscapes by drift analysis, in *Congress on Evolutionary Computation, CEC 2004*, 19–23 June 2004, Portland, OR, USA, 1248–1253. http://dx.doi.org/10.1109/CEC.2004.1331040

Hordijk, W. 1996. A measure of landscapes, *Evolutionary computation* 4(4): 335–360. http://dx.doi.org/10.1162/evco.1996.4.4.335

Jones, T. 1995. *Evolutionary algorithms, fitness landscapes and search*: PhD thesis. The University of New Mexico.

Jones, T.; Forrest, S. 1997. Fitness distance correlation as a measure of problem difficulty for genetic algorithms, in *Proceedings of the Sixth International Conference on Genetic Algorithms*, 15–19 July 1995, University of Pittsburgh, San Francisco, CA, 184–192.

Kauffman, S. 1989. Adaptation on rugged fitness landscapes, in D. L. Stein (Eds.). *Lectures in the science of complexity*. Addison-Wesley, 527–618.

Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. 1983. Optimization by simulated annealing, *Science*, 220(4598): 671–680. http://dx.doi.org/10.1126/science.220.4598.671

Masri, H. 2014. Quantitative economics as a scientific approach to the solution of problems of a complex, *Technological and Economic Development of Economy* 20(3): 590–600. http://dx.doi.org/10.3846/20294913.2014.966350

Merkuryeva, G. 2005. Response surface-based simulation metamodelling methods with applications to optimisation problems, in A. Dolgui, J. Soldek, O. Zaikin (Eds.). *Supply chain optimisation: product/process design, facility location and flow control*. Springer, 205–215. http://dx.doi.org/10.1007/0-387-23581-7_15

Merkuryeva, G.; Bolshakov, V. 2012a. Simulation-based fitness landscape analysis and optimisation for vehicle scheduling problem, in R. Moreno-Díaz, F. Pichler, A. Quesada-Arencibia (Eds.). *EUROCAST 2011, Part I, LNCS* 6927: 280–286. Springer. http://dx.doi.org/10.1007/978-3-642-27549-4_36

Merkuryeva, G.; Bolshakov, V. 2012b. Simulation optimisation and monitoring in tactical and operational planning of deliveries, in *Proceedings of the European Modeling and Simulation Symposium*, 19–21 September 2012, Vienna, Austria, 226–231.

Merkuryeva, G.; Bolshakov, V. 2014. Integrated planning and scheduling built on cluster analysis and simulation optimisation, *International Journal of Simulation and Process Modelling* 9(1–2): 81–91. http://dx.doi.org/10.1504/IJSPM.2014.061450

Merkuryeva, G.; Bolshakov, V. 2015. Integrated solutions for delivery planning and scheduling in distribution centres, in M. Mujica Mota, I. F. De La Mota, D. Guimarans Serrano (Eds.). *Applied simulation and optimization*. Springer, 135–168. http://dx.doi.org/10.1007/978-3-319-15033-8_5

Merkuryeva, G.; Bolshakovs, V. 2010. Vehicle schedule simulation with AnyLogic, in *Proceedings of 12th International Conference on Computer Modelling and Simulation*, 24–26 March 2010, Cambridge, 169–174. http://dx.doi.org/10.1109/UKSIM.2010.38

Merkuryeva, G.; Bolshakovs, V. 2011. Benchmark fitness landscape analysis, *International Journal of Simulation Systems, Science and Technology* 12(2): 38–45.

Merkuryeva, G.; Merkuryev, Y.; Vanmaele, H. 2011. Simulation-Based planning and optimization in multi-echelon supply chains, *Simulation: Transactions of the Society for Modeling and Simulation International* 87(8): 680–695. http://dx.doi.org/10.1177/0037549710366265

Merz, P.; Freisleben, B. 2000. Fitness landscape analysis and memetic algorithms for the quadratic assignment problem, *IEEE Transactions on Evolutionary Computation* 4(4): 337–352. http://dx.doi.org/10.1109/4235.887234

Nagamochi, H.; Ohnishi, T. 2008. Approximating a vehicle scheduling problem with time windows and handling times, *Theoretical Computer Science* 393(1–3): 133–146. http://dx.doi.org/10.1016/j.tcs.2007.12.001

Napalkova, L.; Merkuryeva, G. 2012. Multi-objective stochastic simulation-based optimisation applied to supply chain planning, *Technological and Economic Development of Economy* 18(1): 132–148. http://dx.doi.org/10.3846/20294913.2012.661190

Pereira, F. B.; Tavares, J.; Machado, P.; Costa, E. 2002. GVR: a new genetic representation for the vehicle routing problem, in *Proceedings of the 13th Irish International Conference on Artificial Intelligence and Cognitive* Science *(AICS '02)*, 12–13 September 2002, Limerick, Ireland, 95–102. http://dx.doi.org/10.1007/3-540-45750-x_12

Pitzer, E.; Affenzeller, M. 2012. A comprehensive survey on fitness landscape analysis, in *Recent advances in intelligent engineering systems: studies in computational intelligence,* Vol. 378. Springer, 161–191. http://dx.doi.org/10.1007/978-3-642-23229-9_8

Pitzer, E.; Affenzeller, M.; Beham, A.; Wagner, S. 2011. Comprehensive and automatic fitness landscape analysis using HeuristicLab, in R. Moreno-Díaz, F. Pichler, A. Quesada-Arencibia (Eds.). *EUROCAST 2011, Part I, LNCS* 6927: 424–431. Springer. http://dx.doi.org/10.1007/978-3-642-27549-4_54

Pitzer, E.; Beham, A.; Affenzeller, M. 2012a. Generic hardness estimation using fitness and parameter landscapes applied to Robust Taboo Search and the quadratic assignment problem, in *Proceedings of the 14th International Conference on Genetic and Evolutionary Computation Companion*, 7–11 July 2012, Philadelphia, USA, 393–400. http://dx.doi.org/10.1145/2330784.2330845

Pitzer, E.; Beham, A.; Affenzeller, M. 2013. Automatic algorithm selection for the quadratic assignment problem using fitness landscape analysis, in M. Middendorf, C. Blum (Eds.). *13th European Conference on Evolutionary Computation in Combinatorial Optimization, LNCS*, 3–5 April 2013, Vienna, Austria*,* 7832: 109–120. Springer. http://dx.doi.org/10.1007/978-3-642-37198-1_10

Pitzer, E.; Vonolfen, S.; Beham, A.; Affenzeller, M.; Bolshakov, V.; Merkuryeva, G. 2012b. Structural analysis of vehicle routing problems using general fitness landscape analysis and problem specific measures, in *1st Australian Conference on the Application of Systems Engineering (ACASE'12),* 6–8 February 2012, Sydney, Australia, 36–38.

Rastrigin, L. A. 1974. Extremal control systems, *Theoretical Foundations of Engineering Cybernetics Series* 3. Moscow: Nauka.

Reeves, C. R.; Rowe, J. E. 2002. *Genetic algorithms: principles and perspectives. A guide to GA theory*. Springer.

Reidys, C. M.; Stadler, P. F. 2001. Neutrality in fitness landscapes, *Applied Mathematics and Computation* 117(2–3): 321–350. http://dx.doi.org/10.1016/S0096-3003(99)00166-6

Sakalauskas, L.; Zavadskas, E. K. 2009. Optimization and intelligent decisions, *Technological and Economic Development of Economy* 15(2): 189–196. http://dx.doi.org/10.3846/1392-8619.2009.15.189-196

Schwefel, H. P. 1995. *Evolution and optimum seeking*. Wiley-Interscience.

Sipser, M. 2006. *Introduction to the theory of computation*. 2nd ed. Thomson Course Technology.

Smith, T.; Husbands, P.; Layzell, P.; O'Shea, M. 2002. Fitness landscapes and evolvability, *Evolutionary Computation* 10(1): 1–34. http://dx.doi.org/10.1162/106365602317301754

Stadler, P. F. 2002. Fitness landscapes, in M. Lassig, A.Valleriani (Eds.). *Biological evolution and statistical physics.* Springer, 183–204. http://dx.doi.org/10.1007/3-540-45692-9_10

Vanneschi, L.; Tomassini, M.; Collard, P.; Verel, S. 2006. Negative slope coefficient: a measure to characterize genetic programming fitness landscapes, in *Lecture Notes in Computer Science* 3905: 178–189. Springer. http://dx.doi.org/10.1007/11729976_16

Vassilev, V. K.; Fogarty, T. C.; Miller, J. F. 2000. Information characteristics and the structure of landscapes, *Evolutionary Computation* 8(1): 31–60. http://dx.doi.org/10.1162/106365600568095

Visipkov, V.; Merkuryev, Yu.; Rastrigin, L. 1994. Optimization of discrete system simulation models (Survey), *Automatic Control and Computer Sciences* 28(4): 10–20.

Vonolfen, S.; Affenzeller, M.; Beham, A.; Wagner, S. 2011. Solving large-scale vehicle routing problem instances using an island-model offspring selection genetic algorithm, in *Proceedings of 3rd IEEE International Symposium on Logistics and Industrial Informatics (LINDI)*, 25–27 August, Budapest, Hungary, 27–31. http://dx.doi.org/10.1109/LINDI.2011.6031155

Wagner, S. 2009. *Heuristic optimization software systems – modeling of heuristic optimization algorithms in the heuristiclab software environment*: PhD thesis. Johannes Kepler University.

Weinberger, E. 1990. Correlated and uncorrelated fitness landscapes and how to tell the difference, *Biological Cybernetics* 63(5): 325–336. http://dx.doi.org/10.1007/BF00202749

Galina MERKURYEVA is a Professor at the Institute of Information Technology, Department of Modelling and Simulation of Riga Technical University (Latvia). She earned her Dr sc. ing. degree in 1984 in technical cybernetics and information theory from the Institute of Electronics and Computer Science of the Latvian Academy of Sciences (Latvia) and Dr habil. in 2003 in control of organisational-technical systems from the Institute of Control Sciences of the Russian Academy of Sciences (Russia). Her professional interests and experiences are in the fields of discrete-event simulation, simulation metamodelling, simulation-based optimisation, decision support systems, logistics, production planning and control, supply chain management and simulation-based training. She is the author of more than 170 publications, including 5 books. She is an editor of the Baltic Journal on Sustainability, Technological and Economic Development of Economy.

Vitaly BOLSHAKOV is a senior researcher at the Department of Modelling and Simulation of Riga Technical University, Latvia. In December 2013 he defended his doctoral thesis and received Dr sc. ing. degree. His main research interests and experiences are in the fields of metaheuristic optimisation, fitness landscape analysis and simulation-based optimisation.