SIMULATION-BASED FITNESS LANDSCAPE ANALYSIS AND OPTIMISATION OF COMPLEX SYSTEMS

Vitaly Bolshakov [a]

[a]Riga Technical University, Kalku Street 1, LV-1658, Riga, Latvia

[a]Vitalijs.Bolsakovs@rtu.lv

ABSTRACT

The research is dedicated to the development of methods and algorithms for a simulation-based fitness landscape analysis and optimisation of complex systems. Research is motivated by a wide spread of hard optimization problems nowadays and a relevance of improvement of their solution methods. Application of the fitness landscape analysis methods in the engineering field and tasks of simulation-based optimisation are reviewed and analysed in the paper. A formalised scheme of simulation-based optimisation enhanced by fitness landscape analysis is developed. Strength and weakness of the fitness landscape analysis is researched on the benchmark landscapes, both with relations between efficiency of the optimisation algorithm and structural features of the corresponding fitness landscapes. The prototype of a software tool for the fitness landscape analysis of simulation optimisation problems is developed. Approbation of the developed methods is performed in optimisation of vehicle schedule and routes in the delivery planning task. Various metaheuristic optimisation scenarios with application of the fitness landscape analysis are investigated.

Keywords: simulation, optimisation, fitness landscape analysis, metaheuristic methods, vehicle scheduling

1. INTRODUCTION

In various cases, 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 for such problems due to a high number of parameters of an optimised system, existence of stochastic parameters and a large solution 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 an optimisation problem can be performed. At the present time, simulation optimisation technology is a necessary tool in optimisation of complex systems, where solution evaluation can be complicated. Simulation-based fitness landscape analysis provides an efficient approach to analysis of suitability of the optimisation algorithms.

Nowadays, fitness landscape analysis methods are used for the determination of the problem hardness for the metaheuristic algorithms (Stadler et al. 2002, Pitzer and Affenzeller 2012). However, there are few researches on fitness landscape analysis within simulation optimisation of complex systems. Simulation-based fitness landscape analysis will allow better selection of optimisation algorithms, as well as allowing for construction and adjustment of the most appropriate algorithm. The research is aimed at developing methods for the simulation-based fitness landscape analysis and optimisation of complex systems.

The structure of the paper is as follows. Section 2 gives formal definitions of fitness landscape and its analysis techniques. The problem of the simulation-based fitness landscape analysis is defined. Section 3 discusses the landscapes of the benchmark problems. Section 4 presents a procedure of the simulation-based fitness landscape analysis. A case study problem is given. Section 5 describes application of metaheuristic optimisation methods in solving of a combined vehicle routing and scheduling task.

2. LITERATURE REVIEW AND PROBLEM STATEMENT

2.1. Simulation-based Optimisation for NP-hard Problems

Modern optimisation problems in logistics and industry are characterized by large dimensions, uncertainty and nonlinearity. A factor that strongly influences the hardness of the optimisation problem is computational complexity of the problem. The research focuses on the NP-hard problems. Other factors that strongly influence the hardness of the optimisation problem can be the stochastic nature of the optimised system and the hardness of obtaining the analytical form of the objective function. To find solutions of such complex, large-scale, stochastic optimisation problems simulation-based optimisation is applied.

Numerical optimisation methods form a natural choice in solving complex stochastic optimisation problems, where the closed form of the objective function is frequently unknown (Gosavi 2003). These methods include metaheuristic optimisation methods, which facilitate finding good solutions to large and complex optimisation problems in a reasonable time with the application of different heuristic and stochastic methods. Although metaheuristic methods don't guarantee to find the optimal solution of the problem, there is a high interest for such methods in the applied optimisation of real life problems (Glover and Kochenberger 2003). These methods include such examples as the *Genetic Algorithm* (GA) (Goldberg 1989) and *Evolution Strategy* (ES) (Schwefel 1995). The application of the metaheuristic and other numerical methods becomes more important for especially hard optimisation problems such as NP-hard combinatorial optimisation problems (Dreo et al. 2006).

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

2.2. The Concept of Fitness Landscape Analysis

The 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 systems, and it is widely considered in literature (Weinberger 1990, Jones and Forest 1997, Stadler 2002). The fitness landscape is interpreted as a combination of a fitness function of the optimisation problem and the relationships or a distance metric between the solutions in the search space (Reeves and Rowe 2002).

It was proposed that the structures of a fitness landscape affect the way, in which a search space is examined by a metaheuristic optimisation algorithm. The 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 and Rowe 2002). With the landscape analysis it is possible to get measures of the problem's difficulty, and the recommended configuration of an optimisation algorithm. Searching for better problem subclass specific algorithms and configurations will provide useful knowledge on the problem solution scenarios (Pitzer and Affenzeller 2012).

Formal definitions of fitness landscapes are provided in the literature. In the following definition (Jones 1995) a *representation space* $\boldsymbol{R}$ defines a set of representations and the *search operator* ϕ is defined as a function $\phi: \boldsymbol{M}(\boldsymbol{R}) \times \boldsymbol{M}(\boldsymbol{R}) \rightarrow [0, 1]$, where $\boldsymbol{M}(\boldsymbol{R})$ is a multiset of representations. A value of $\phi(v, w) = p$ for $v, w \in \boldsymbol{M}(\boldsymbol{R})$ defines a probability p that v will be modified to w by application of the operator ϕ. The fitness landscape is defined as the 5-tuple:

$$\boldsymbol{L} = (\boldsymbol{R}, \phi, f, \boldsymbol{F}, >_{\boldsymbol{F}}), \quad (1)$$

where f is a fitness function; $\boldsymbol{F}$ is the fitness space with a partial order $>_{\boldsymbol{F}}$. The landscape can be represented as a directed labelled graph $G_{\boldsymbol{L}} = (V, E)$, where vertices are $V \subseteq \boldsymbol{M}(\boldsymbol{R})$, and edges are $E \subseteq V \times V$. In this representation, a vertex $v \in V$ is labelled as $f(v)$, and edge (v, w) is labelled $\phi(v, w)$. Similar to structures of nature landscapes hill ridges, valleys and other structures can be identified in the fitness landscape. Following structures are formalised in the literature: *peak* (or *maximum*), *global-optimum*, *local-optimum*, *plateau*, *basin of attraction* (Jones 1995).

The major highlighted factors, which affect the hardness of the optimisation problem, are: the m*odality*, which defines a number and density of optima in a search space (Reeves and Rowe 2002); *ruggedness* that characterizes the impact of all landscape structures on the hardness of the search (Merz and Freisleben 2000); and *neutrality*, which characterizes a number of plateaus (Reidys and Stadler 1998).

2.3. Fitness Landscape Analysis Techniques

Different techniques have been developed for a fitness landscape analysis by evaluating its structural characteristics (Jones and Forest 1997, Vassilev et al. 2000, Smith et al. 2002, Collard et al. 2004).

Fitness landscape analysis techniques apply different strategies for data collection based on simple moves, which generate a trajectory through the landscape. In the *Random Walk*, a solution candidate is randomly modified repeatedly. In the *Adaptive Walk*, a certain number of mutations are performed to generate a set of neighbours, and then the best one is selected from this set (Kauffman 1989). The *Up-Down Walk* is similar to the adaptive walk, but the direction of the walk is reversed when a local optimum is reached (Vassilev et al. 2000). *Neutral Walks* explore "flat" areas (Reidys and Stadler 1998).

The *statistical analysis* proposed by Weinberger (1990), calculates the autocorrelation function in the random walk to measure the ruggedness of the landscape. In case of a high correlation between fitness values the landscape is considered less rugged. In the first step a time series of fitness values $\{f_t\}^N_{t=1}$ is obtained in a landscape walk of N moves. Then, an autocorrelation function $\rho(\Gamma)$ is calculated:

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

where $E(f_t)$ the expectation and $V(f_t)$ is the variance of a sequence $\{f_t\}^N_{t=1}$. For smooth landscapes the autocorrelation of a random walk is close to 1 and tends to zero for rugged (Reeves and Rowe 2002). Another statistical measure is correlation length, which defines a distance beyond which two sets of fitness points becomes uncorrelated. A longer correlation length indicates a smooth landscape.

The ***information analysis*** interprets a fitness landscape as an ensemble of objects, which are characterized by their form, size and distribution. These objects consist of a point in the fitness landscape and the nearest neighbours of this point. The information analysis is based on the information theory, and four information measures are proposed by Vassilev et al.

(2000). The *information content* $H(\varepsilon)$ is a measure of entropy in the system. In case of high information content, the landscape is more rugged. *Partial information content* $M(\varepsilon)$ characterizes the modality of the obtained fitness string. The *information stability* ε^* characterizes a magnitude of optimums in the obtained landscape fitness path. The *density-basin information* $h(\varepsilon)$ analyses the variety of flat and smooth sections in the landscape. Information measures are calculated with notice to a calculation accuracy which is defined by a parameter ε, which defines a threshold of slopes in the fitness path (Vassilev et al. 2000).

2.4. Problem Setup of the Simulation-based Fitness Landscape Analysis and Optimisation

To extend the concept of the fitness landscape analysis for its application in simulation-based optimisation, the concepts of the simulation-based fitness landscape analysis are introduced in (Bolshakov 2013). The formal definition of the simulation fitness landscape $\boldsymbol{L'}$ with an assumption that a simulation model provides real value output is:

$$\boldsymbol{L'} = (\boldsymbol{R}, \phi, S), \tag{3}$$

where $\boldsymbol{R}$ is a representation space, ϕ is a search operator and S is a simulation model with one output variable.

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

Landscape walk module LW can be interpreted as follows:

$$\vec{x}_{t+1} = LW(\vec{x}_t, \phi, \hat{y}), \tag{4}$$

where t is a number of completed walk iterations, $\vec{x}_{t+1}$ is a vector of simulation model input variables for a current iteration, $\vec{x}_t$ are input variables at the previous iteration and $\hat{y}$ is an output of a simulation model. Output of the landscape walk module is a vector $\vec{x}$ = $(x_1, x_2, \ldots, x_k)$, $\vec{x} \in \boldsymbol{R}$.

Simulation model S evaluates the performance of a system. Its output is estimated by $\hat{y} = \mathrm{E}[y]$, where $y \in \mathbb{R}$ is simulation output in each replication and $\mathrm{E}[\cdot]$ is the mathematical expectation. As a result of process integration of modules LW and S, a number of time series $\{\hat{y}_t\}^N_{t=1}$ are generated.

The module of **statistical and information analysis** performs analysis of fitness values time series, and calculates the landscape statistical and information analysis measures. A set of measure values is obtained for different values of Γ and ε.

The module of **construction and tuning of an optimisation algorithm** allows selecting the appropriate optimisation algorithm and adjust its parameters for optimisation of a complex system, which is simulated by S. Selection of the algorithm, its components and parameters is based on the data from simulation-based fitness landscape analysis and the data on benchmark landscapes. The module output defines the selected metaheuristic optimisation algorithm and its configuration: the representation $\boldsymbol{R}$ and a set Φ of search operators which form the optimisation algorithm. The selection of the algorithm and its configuration is based on the rules and recommendations which are applicable for known values of the landscape measures.

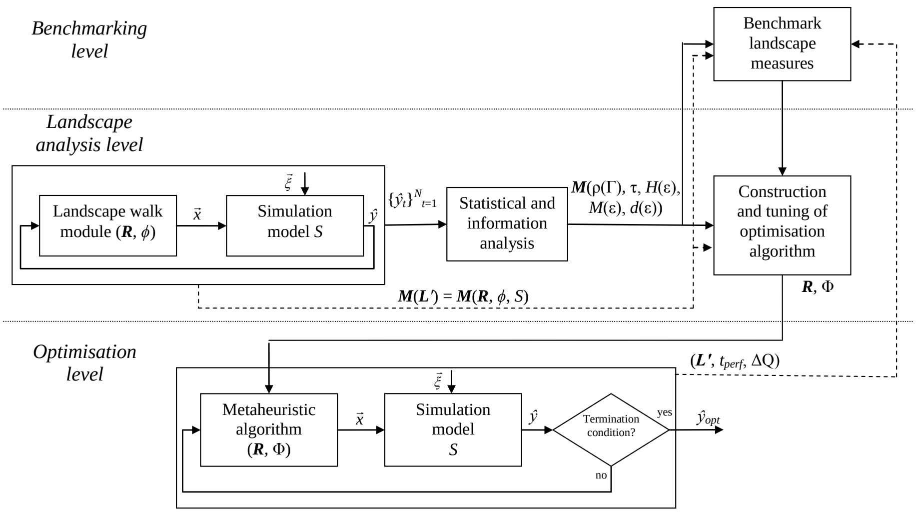

Figure 1: Simulation-based Optimisation with Fitness Landscape Analysis

Proceedings of the Int. Conf. on Harbor Maritime and Multimodal Logistics M&S, 2014
ISBN 978-88-97999-39-3; Bruzzone, Del Rio Vilas, Longo, Merkuryev, Piera Eds

At the *optimisation level*, the **metaheuristic algorithm** uses the representation method and a set Φ of suggested operators to optimise the problem. The **termination condition** determines whether the suitable solution is found. When the optimisation cycle is terminated, the best found solution $\hat{y}_{opt} = \langle \vec{x}, \hat{y} \rangle$ is selected.

3. BENCHMARK FITNESS LANDSCAPE EXPERIMENTAL ANALYSIS

3.1. Experimental Analysis of Benchmark Landscapes

The following four fitness functions are widely used for benchmarking of genetic algorithms and were selected for estimating and analysing statistical and information measures of benchmark fitness landscapes. They are Sphere and Rosenbrock functions (De Jong 1975), Rastrigin function (Rastrigin 1974) and Ackley function (Ackley 1987). Experiments with the benchmark landscapes were performed with same number of variables and within a same search domain. Eight different fitness landscapes as a combination of four different benchmark functions and two types of representations are analysed. For a detailed analysis of benchmark fitness landscapes, a software prototype in Java was developed and applied. To estimate structural measures of these landscapes, three series of landscape analysis experiments were performed (Merkuryeva and Bolshakovs 2011).

In the first series of experiments it was found that while correlation measures show dependence on the length of the path generated by a random walk, the behaviour of information content measures does not demonstrate this effect.

In the second series of experiments, the autocorrelation for different benchmark landscapes and lags was defined for two types of solution representation. Correlograms obtained for real-value and binary coded benchmark landscapes show the higher autocorrelation for real-value coded fitness landscapes that make search processes easier in practice.

In the third series of experiments, different information measures for all benchmark landscapes and different ε values were estimated. At ε = 0 information measures become almost identical and essentially do not provide a new information about structures of specific fitness landscapes. At the same time, smaller values of the information content for the Rosenbrock function compared to the Sphere indicate the higher degree of flatness with respect to rugged areas of the landscape (Merkuryeva and Bolshakovs 2011).

3.2. GA Optimisation Experiments with Benchmark Fitness Functions

To find the correlation between the results of fitness landscape analysis and hardness of a real problem for an evolutionary algorithm, a series of optimisation experiments were performed with benchmark landscapes (Merkuryeva and Bolshakovs 2011). GA with one point crossover and corresponding mutation operator was used to estimate a cumulative probability of success for different benchmark landscapes. The results of optimisation experiments show that in most cases, except for the Rosenbrock function, GA found solutions on real-value coded benchmark landscapes are better than on the binary ones that was predicted within the statistical analysis. As the autocorrelation between neighbourhood fitness points is high, it is easier for the genetic algorithm to move to a point with better fitness.

4. SIMULATION-BASED FITNESS LANDSCAPE ANALYSIS

4.1. Fitness Landscape Analysis Tool

A procedure for the simulation-based fitness landscape analysis in a prototype of an analysis tool contains three following stages (see Fig. 2):

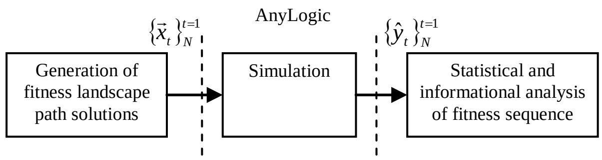

Figure 2: Main Stages of Fitness Landscape Analysis

The procedure is as follows. A developed Java application performs a random walk on the problem fitness landscape with application of mutation operator and produces a sequence $\{\vec{x}_t\}_N^{t=1}$ of landscape path candidate solutions. The solutions are evaluated by simulation model of the analysed system in AnyLogic *parameter variation* experiment with different vectors of input parameters, which are defined in the obtained trajectory. As a result, the model generates an array of fitness values $\{\hat{y}_t\}_N^{t=1}$. Finally, calculation of statistical and information fitness landscape measures is performed on a sequence of obtained fitness values.

In statistical analysis, correlation length and values of autocorrelation function for different lags are calculated. In the information analysis at first a value of information stability ε^* is determined. Then iterative calculations of information content, partial information content and density-basin information are performed for different values of ε within interval [0, ε^*).

Experimental data from several random walks is collected during analysis. As all random walks are started at different random positions, fitness landscape measures are obtained for a large part of the landscape.

4.2. Case Study

A case study of the vehicle scheduling problem with time windows (VSPTW) is considered in the research. Vehicles with various parameters deliver different types of goods from one distribution centre to various stores. A sequence of stores in a route, moving, loading and unloading times are defined for each trip. Delivery time windows and average demand are defined for each store. Vehicle capacities are limited and known. The problem is aimed at assigning vehicles to trips in order to minimise the total idle time of all vehicles, which is defined as a sum of time periods when a vehicle is waiting for the next trip. Vehicle scheduling problem (VSP) is frequently reviewed in the studies (Eiiyi et al. 2008, Nagamochi and Ohnishi 2008). In practice, VSP can also be complicated by stochastic processes, and this is a reason to apply simulation optimisation to solve such problems.

The following two sets of decision variables are introduced: v_i and t_i, where v_i is a vehicle assigned and t_i is a start time of the trip i. The problem constraints are vehicle capacity constraints and delivery time windows of stores.

The objective function f is specified as follows:

$$f = \sum_{i=1}^{N} T_{idle}^{i} + k_1 T_c + k_2 T_m + k_3 T_0 + k_4 N_{ol} + k_5 N_{ot} \rightarrow \min, \quad (5)$$

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 have overdone 24 working hours; and N_{ot} is a number of vehicles that are overloaded. All indexes for unsatisfied constraints are multiplied with penalty coefficients k_i.

To determine the fitness of potential vehicle schedule solutions, discrete event simulation model in AnyLogic is developed (Merkuryeva and Bolshakovs 2010). The simulation model evaluates the efficiency of a potential vehicle schedule by estimating the total idle time of all vehicles. To validate the model the existing schedule of a case study was simulated. In the experimental analysis, when the vehicle moving time between two route points was defined as a random variable with normal distribution it was determined that stochastic nature of the vehicle moving time has an impact on the vehicle idle time which grows with the growth of variance. At same time, a sum of all moving intervals for a vehicle is not affected by variance of moving times.

To solve the vehicle scheduling problem with time windows, three optimisation scenarios are defined in the research:

1. Optimisation in OptQuest optimisation tool.
2. Simulation-based fitness landscape analysis and optimisation of the problem in the developed prototype.
3. Fitness landscape analysis and optimisation of the problem in the HeuristicLab framework.

As the simulation model of the vehicle schedule is developed in the AnyLogic software, optimisation tool OptQuest was applied. But, in experiments OptQuest was not able to obtain good solutions of the VSPTW, thus this scenario is not described in the paper.

4.3. Problem Research with Developed Simulation-based Tools

In this scenario, the VSPTW is sequentially analysed by the simulation-based fitness landscape analysis tool prototype and optimised in simulation-based optimisation by tuned genetic algorithm.

Here, solution of the VSP is encoded as an integer vector chromosome, which length is twice the number of trips. Genes with even numbers represent start times of corresponding trips, and odd genes define the assigned vehicle for this trip. To perform the random walk on the fitness landscape, a mutation operator is introduced that changes one randomly selected trip in the solution candidate. For the selected trip a new randomly chosen vehicle is assigned, and start time is shifted by a certain constant value.

Information and statistical measures of the VSPTW fitness landscapes with stochastic and deterministic input data received in experiments are given in Table 1.

Information measures demonstrate that the landscape of problem with stochastic data has higher entropy and should have higher modality. According to the landscape measures, problem with stochastic data should be more complex for the optimisation algorithm as values of autocorrelation function between neighbour solutions ρ(1) are lower.

In both cases the information content is relatively high, and fitness landscape of the optimisation problem is relatively rugged. The partial information content is low, and as a result, the modality of fitness landscape is low. The results of the fitness landscape analysis lead to a conclusion that the case study problem is not hard for evolutionary algorithms. Comparative analysis shows that landscape of VSPTW is less rugged than landscapes of benchmark fitness functions whose solutions are coded in binary chromosomes. Thus the analysed problem could be solved with the GA no worse, than mentioned benchmark problems.

Table 1: Information and Statistical Measures

Model input data	$H(0.1)$	$M(0.1)$	$h(0.1)$	ε*	ρ(1)	ρ(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

In the simulation optimisation, the GA is applied. The optimisation tool is implemented as a Java class, which interacts with the simulation model via ‘Parameter variation’ experiment in AnyLogic. In experiments with population size 200, one point crossover operator for data encoded in real numbers and a described mutation operator the best found solution allowed decreasing the total idle time comparing with the original schedule.

In series of optimisation experiments, simulation model with deterministic data is used and termination condition is set to occur when a large number of generations are generated without improvement of the best solution in the population. Genetic operators are customized for operating with the proposed structure of the chromosome: one point crossover operator for real vectors and above described mutation operator are applied. As fitness evaluation with simulation is time consuming, caching of fitness values is applied. If the solution was already evaluated its fitness value can be returned without simulation from an array of stored fitness values. Optimisation results show that a solution which satisfies all constraints can be found. Acceptable results are obtained with the population size higher than 1000 chromosomes (Merkuryeva and Bolshakov 2012a). But the optimisation algorithm needs improvements, as many found solutions of large dimension instances do not satisfy part of soft constraints

4.4. Fitness Landscape Analysis and Optimisation in HeuristicLab

To perform a faster and more comprehensive analysis, the simulation model was reimplemented as a plug-in of HeuristicLab (Wagner 2009) maintaining all its logic. To enhance the quality of optimisation results, permutation encoding for the VSP 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 the 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 the landscape analysis experiments is created to compare values between different landscapes:

1. Comparison of different mutation operators.
2. Impact measurements of stochastic variables during simulation.
3. Comparison between existing and proposed encodings.

Full results of comprehensive analysis experiments are presented in Bolshakov et al. (2011). For the integer vector representation, fitness landscapes of two operators are analysed. The single position replacement manipulator (*VSPManipulator*) changes the start time of the trip to a new uniformly distributed random number, but the single position shift manipulator (*VSPShiftManipulator*) shifts the start time with a uniformly distributed random number. In random walk, values of autocorrelation function are slightly lower for the replacement operator. In up-down walk the situation is the opposite: replacement mutation has higher correlation than shift mutation, but the three artificial problems are different to the others (see Fig. 3; black dots are for *replacement* and green for *shift* mutator). It can be concluded that for VSPs that the main impact on the local landscape structures has a number and variety of trips.

The plug-in was supplemented with additional logic to estimate the affect of simulation model's stochastic variables on the landscape measures. In following series of experiments vehicle movement times between customers are shifted by a random number that has symmetric triangle distribution in the interval [-20, +20] minutes. The autocorrelation value ρ(1) is lower for landscapes of noisy problems in these experiments. The addition of similar noise has different impact on different instances, which can be measured by $H(0)$ or $M(0)$ in random walks, which values are higher for landscapes with noise. It is determined in experiments, that higher number of replication reduces the impact of the noise on the information measures, although no significant difference of correlation length and autocorrelation values between different numbers of replications was found. The information content's $H(0)$ value is higher for the problem instances with additional noise, especially when only one replication is used.

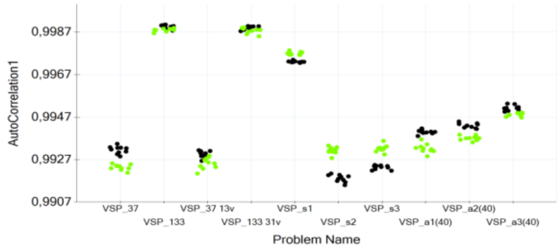

Figure 3: Autocorrelation in Up-Down Walks

To compare the fitness landscape analysis measures between different VSP representations, experiments for each encoding are performed. The value of the autocorrelation function in random and up-down walks is lower for the permutation encoding, which means that landscapes of this encoding should be more rugged.

Evolution Strategy (ES), Simulated Annealing (SA) and Genetic Algorithm were applied in the comparison of VSP optimisation results. For integer encoding, both ES and SA algorithms are fast and highly successful, and it is possible to find solutions with better quality with ES (Fig. 4). GA finds even better solutions, but requires a higher number of evaluations. The permutation encoding is found to be more effective in optimisation of the VSP, as it reduces the search space, even though the fitness landscape for the permutation encoding has to be more rugged. Although it is also found, that for large dimension instances factor of landscape ruggedness dominates the reduction of the size of search space.

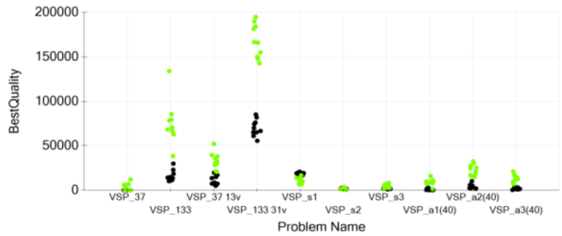

Figure 4: Quality of Best Found Solutions with ES

It is experimentally found that the statistical analysis can predict the performance of these operators. For the GA instances, where the shift operator is better, the autocorrelation for this operator also is higher (Fig. 3 and 4; black dots are for *replacement* and green for *shift* mutator). The same dependency was also found for the ES algorithm.

It is concluded, that optimisation using the ES is the best choice for the solution of the vehicle scheduling problem with time windows. If the GA is selected as the optimisation algorithm, permutation encoding has to be chosen unless the problem contains more than 100 trips. In the integer vector encoding, selection of the appropriate mutation operator is based on the landscape analysis measures: an operator, which has the highest autocorrelation value in the up-down walk, should be selected.

5. APPLICATION IN PRODUCT DELIVERY PLANNING

Two combined optimisation tasks of the integrated delivery planning proposed in Merkuryeva and Bolshakov (2012b) are solved in this section. General task is the determination of the best routes and schedule for the vehicles to deliver goods from a distribution centre to the stores. Optimal distribution of routes and vehicles should minimise a number of used vehicles, total delivery distances, with minimisation of vehicle idle times. The route and schedule plan must fulfil the constraints, such as capacities of vehicles, time windows and warehouse capabilities. For the solution of this task, two optimisation problems are solved sequentially. First task is solved as a vehicle routing problem with time windows. For the second task a route scheduling problem statement is defined that aims optimisation of a schedule of predefined routes.

5.1. Vehicle Routing Experiments

The classical flow-based vehicle routing problem's with time windows (VRTPW) statement (Cordeau et al. 2001) was applied in the research. All optimisation experiments were performed with application of the Island Offspring Selection Genetic Algorithm (Affenzeller et al. 2009), which is a special type of genetic algorithm that combines features of coarse-grained parallel GA and GA with offspring selection. Experiments were performed with HeuristicLab optimisation framework (Wagner 2009). A GVR crossover was selected as best for the VRPTWs of the case study, as it works with an unlimited number of vehicles, but provides best results in terms of keeping routes not overloaded.

It was found for the considered case, that in best found solutions many routes are 1 to 3 customers long. A number of stores is limited due to the small capacity of the vehicles, and not because of short time windows.

5.2. Vehicle Route Scheduling Problem Statement

It is assumed in the definition of the classical VRPTW, that any vehicle may perform only one route in the planning horizon. In the investigated business case, all routes are shortened by the capacity of vehicles, which leads to the ineffective solutions of the vehicle routing problem. To overcome these obstacles, the route scheduling problem is introduced. It can be formulated on a basis of the VSPTW. In the formulated problem, the routes correspond to the trips in the VSPTW and vehicles may perform any fair number of routes during the day. As far as the final solution of the VRPTW task should be feasible for the capacity and time window constraints, it can be optimised by combining and compacting routes to increase a vehicle utilisation. Application of the vehicle scheduling for the solution of vehicle routing problem allows reducing a number of required vehicles.

A full formal statement of the route scheduling problem is described in (Merkuryeva and Bolshakov 2012b). The decision variables are ones introduced in the routing model (Cordeau et al. 2001), i.e., sets x and s, except that $x_{ijk} = 1$ states that for vehicle k route j will be the next after route i. Two types of soft constraints are introduced: 1) time window constraints; 2) overtime constraints. A fitness function f of the route scheduling problem summarizes all idle times and a number of constraint violations multiplied by penalty values:

$$f = \sum_{k \in V} l_k + p_{ad} N_{ad} + p_{ot} N_{ot} \to \min, \qquad (6)$$

where l_k is the total idle time of a vehicle k; V is a set of available vehicles; N_{ad} is a number of vehicles, which leave customer after due time; N_{ot} is a number of vehicles, which are scheduled to work with overtime; p_{ad} and p_{ot} are the penalty values for late deliveries and vehicle overtimes, correspondingly.

5.3. Vehicle Route Scheduling Experiments

To resolve the vehicle route scheduling problem, a plug-in in HeuristicLab optimisation framework is developed. In the plug-in, fitness function (6) evaluator simulates a schedule of a solution candidate. A permutation encoding of the VSPTW is applied for the route scheduling, but the trips here are represented by the routes.

Several series of optimisation experiments were performed to determine a suitable algorithm for the route scheduling and numbers of solution evaluations to obtain candidate solutions of the equal fitness are compared. Following algorithms were examined: ES, GA, Island GA with 5 islands and Offspring Selection GA (Affenzeller et al. 2009). The ES was chosen as most suitable, for its ability to provide the best found optimal results of the route scheduling with fewer evaluations (Merkuryeva and Bolshakov 2012b).

A sample experiment based on one day plan and specific demand data for 53 stores is described. The best found solution obtained by the IOSGA for the VRPTW defines 34 routes (Fig. 5). Here, it is possible to combine these routes due to the long time windows. The ES (20+100) algorithm was applied for the route scheduling problem which input data is based on the considered VRPTW solution. As a result, the optimal scheduling solutions were found with all constraints satisfied if at least 6 vehicles are available (see Fig. 5).

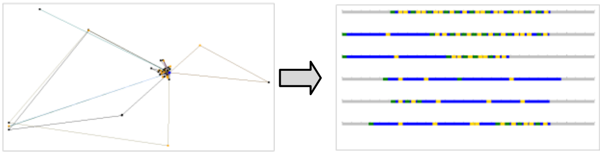

Figure 5: VRPTW and Scheduling Solution of Sample Instance

The proposed vehicle scheduling method that complements vehicle routing can be applied in the vehicle routing and scheduling, when routes are short in comparison with a planning horizon.

6. RESULTS AND CONCLUSIONS

Review of formal definitions of fitness landscape and its structures, with the review of fitness landscape analysis methods allowed development of the formal scheme for the simulation-based optimisation, enhanced with the fitness landscape analysis. Experimental fitness landscape analysis of benchmark landscapes allowed finding the relations and dependencies between structural features of benchmark fitness landscapes, their measures and behaviour of optimisation algorithm on these landscapes.

The developed simulation-based fitness landscape analysis procedure allowed implementation of a software tool prototype for fitness landscape analysis. Application of this tool provided analysis of the vehicle scheduling problem with time windows in simulation-based optimisation. The comprehensive experimental fitness landscape analysis of this problem allowed determination of problem specific properties and internal characteristics of problem's fitness landscape, which, provided development of recommendations for optimisation scenarios of the vehicle scheduling problem with time windows. Experimental results show that it is possible to use fitness landscape analysis for enhanced optimization of applied problems, but with notice to the stochastic data of simulation optimisation.

The developed methods were applied in the solution of delivery planning operational level optimisation tasks, which allowed improving the overall solutions of vehicle routing and scheduling problem with time windows.

ACKNOWLEDGMENTS

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., 1987. An empirical study of bit vector function optimization. *Genetic Algorithms and Simulated Annealing*, 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.

Bolshakov, V., 2013. *Simulation-based Fitness Landscape Analysis and Optimisation of Complex Systems.* Thesis (PhD). Riga Technical University.

Bolshakov, V., Pitzer, E., Affenzeller, M., 2011. Fitness Landscape Analysis of Simulation Optimisation Problems with HeuristicLab. *Proceedings of the UKSim 5th European Symposium on Computer Modeling and Simulation*, pp 107-112, 16.-18. November, Madrid (Spain).

Beerenwinkel, N., Pachter, L, Sturmfels, B., 2007. Epistasis and Shapes of Fitness Landscapes. *Statistica Sinica,* 17(4), 1317-1342.

Collard, P., Verel, S., Clergue, M., 2004. Local Search Heuristics: Fitness Cloud versus Fitness Landscape. *Proc. of 16th European Conference on Artificial Intelligence,* pp. 973-974.

Cordeau, F., Desaulniers, G., Desrosiers, J., Solomon, M.M., Soumis, F., 2001. VRP with Time Windows In: Toth P.m Vigo D. eds. *The vehicle routing problem.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 157-193.

De Jong, K. D., 1975. *An analysis of the behavior of a class of genetic adaptive systems*. Thesis (PhD). 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.

Eiiyi, 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.

Glover, F., Kochenberger, G.A., 2003. *Handbook of metaheuristics. International Series in Operations Research & Management Science*: Springer.

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.

Jones, T., 1995. *Evolutionary Algorithms, Fitness Landscapes and Search.* Thesis (PhD). The University of New Mexico.

Jones, T., Forrest, S., 1997. Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithms. *Proceedings of the Sixth International Conference on Genetic Algorithms.* pp.184-192, San Francisco, CA.

Kauffman, S., 1989. Adaptation on rugged fitness landscapes. *Lectures in the Science of Complexity*: Addison-Wesley, 527-618.

Merkuryeva, G., Bolshakovs, V., 2010. Vehicle Schedule Simulation with AnyLogic. *Proceedings* *of 12th International Conference on Computer Modelling and Simulation*, pp. 169-174, Cambridge.

Merkuryeva, G., Bolshakovs, V., 2011. Benchmark Fitness Landscape Analysis. *International Journal of Simulation Systems, Science and Technology*, 12(2), 38-45.

Merkuryeva, G., Bolshakov, V., 2012a. Simulation-based fitness landscape analysis and optimisation for vehicle scheduling problem. In: *EUROCAST 2011, Part I, LNCS 6927*: Springer, 280-286.

Merkuryeva, G., Bolshakov, V., 2012b. Simulation Optimisation and Monitoring in Tactical and Operational Planning of Deliveries. *Proceedings of the European Modeling and Simulation Symposium*, pp 226-231.

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.

Nagamochi, H., Ohnishi, T., 2008. Approximating a vehicle scheduling problem with time windows and handling times *Theoretical Computer Science*, 393(1-3), 133-146.

Pitzer, E., Affenzeller, M., 2012. A Comprehensive Survey on Fitness Landscape Analysis. In: *Recent Advances in Intelligent Engineering Systems*: Springer, 161–191.

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., 1998. Neutrality in Fitness Landscapes. *Applied Math. & Comput.*, 117, 321-350.

Schwefel, H.P., 1995. *Evolution and Optimum Seeking*: Wiley-Interscience.

Smith, T., Husbands, P., Layzell, P., O’Shea, M., 2002. Fitness landscapes and evolvability. *Evolutionary Computation*, 10(1), 1-34.

Stadler, P.F., 2002. Fitness Landscapes. *Applied Math. & Comput.*, 117, 187-207.

Vassilev, V.K., Fogarty, T.C., Miller, J.F., 2000. Information Characteristics and the Structure of Landscapes. *Evolutionary Computation*, 8(1), 31-60.

Wagner, S., 2009. *Heuristic Optimization Software Systems - Modeling of Heuristic Optimization Algorithms in the HeuristicLab Software Environment.* Thesis (PhD). Johannes Kepler University.

Weinberger, E., 1990. Correlated and Uncorrelated Fitness Landscapes and How to Tell the Difference. *Biological Cybernetics*, 63(5), 325-336.

AUTHORS BIOGRAPHY

Vitaly Bolshakov is a 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. His main research interests and experiences are in the fields of metaheuristic optimisation, fitness landscape analysis and simulation-based optimisation.