AUTHOR, OLDDEGREE

Evaluation of a Genetic Algorithm on generating critical Scenarios in a Traffic Simulation

Master's Thesis

to achieve the university degree of

Master of Science

Master's degree programme: Telematik

submitted to

Graz University of Technology

Supervisor

Dr. Some Body

Institute for Softwaretechnology
Head: Univ.-Prof. Dipl-Ing. Dr.techn. Some One

Graz, November 2013

This document is set in Palatino, compiled with pdfLaTeX2e and `Biber`.

The LaTeX template from Karl Voit is based on KOMA script and can be found online: https://github.com/novoid/LaTeX-KOMA-template

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis.

Date

Signature

Abstract

This is a placeholder for the abstract. It summarizes the whole thesis to give a very short overview. Usually, this the abstract is written when the whole thesis text is finished.

Contents

List of Figures

1 Introduction

This Thesis will use a Genetic Algorithm in order to generate critical Driving Scenarios for testing ADAS/AD Functionality in vehicles. While generating these scenarios is the objective, the main task of the thesis will evolve around the implementation of the Genetic Algorithm as well as the Optimization of its Hyperparameter.

1.1 Research Questions

1.1.1 Research Question 1

Is a Genetic Algorithm suitable for generating critical driving scenarios compared to a random generation?

1.1.2 Research Question 2

Is it possible to improve the performance of a Genetic Algorithm using Taguchi Orthogonal Array Testing?

Taguchi Orthogonal Arrays allow for minimal experiment runs when testing across different settings Roy, 1990). It will be analyzed if this method can be used to

1.1.3 Research Question 3

Can a hypertuned Genetic Algorithm generalize on different start scenarios?

Due to performance considerations, only one start scenario was defined. When evaluating the optimzied Genetic Algorithm, its performance will be compared across different start scenarios.

1.2 Shortcomings

This Master Thesis started with the developement of the Traffic Manger and thus progress was closely linked. Without a working simulations, no genetic alogirthmis could be tested. Due to time and performance constraints, it is not possible to test a full driving stack like autoware, as well as other professional ADAS/AD functions. In this Thesis, internal functions like Time-To-Collision and Emergency Braking will be optimized. The learned information on e.g. optimal hyperparameter settings can then be applied in further steps to test these functions. This will however not be tackled by this thesis.

Performance is also a problem and will lead to many shortcuts that need to be taken. There is a hughe number of possible compations of hyperparamter, so only a handful can be tested. In further chapers, these shortcuts will be explained and their relevancy will be dicussed.

2 Foundations

2.1 Genetic Algorithm

General Pioneered by John Holland (1974) , Genetic Algorithms (GAs) imitate natural selection and the Darwinian principle called "survival of the fittest". Genetic Algorithms search the solution space using a population of individuals. Each individual contains one chromosome, which serves as a candidate solution. Using genetic operations, the individuals mate among themselves and mutate independently. Which individual is chosen depends on a fitness function, which defines the search problem. Individuals that do poorly on the given problem die out, while "strong" individuals propagate. Genetic Algorithms iteratively (each iteration is call "generation") optimize by manipulating a population of potential solutions (Srinivas and Patnaik, 1994). Chromosomes consist of individual genes, which form a solution to the search problem. Over generations using genetic operations, gene values and position will be optimized, resulting in interativly better solutions (Srinivas and Patnaik, 1994). This masters thesis will utilize generational genetic algorithms, where each generation, the entire population is replaces. Contrary to that, steady state genetic algorithms only replace a small fraction of the population at a time (Srinivas and Patnaik, 1994).

Find paper

Genetic algorithms are a subclass of evolutionary algorithms (Mills, Filliben, and Haines, 2015).

Pros This search method on the basis of biological principles allows trough its population for a global and dispersed search which avoids various shortcommings of local search techniques (Grefenstette, 1986). Especially on difficult search spaces with lots of local optima, a genetic alogorithm is less

prone to get stuck on a substandard solution (Katoch, Chauhan, and Kumar, 2021, Xia et al., 2019, Majumdar and Ghosh, 2015). Genetic Algorithms offer advantages for complex optimization and non-deterministic polynomial (NP-hard) problems (Hussain and Muhammad, 2020).

“GAs can find good solutions within a large, ill-defined search space, and can be readily adapted to a wide variety of search problems (Mitchell, 1998)” (Mills, Filliben, and Haines, 2015)

Grefenstette, 1986 further emphasis its ability to outperform gradient techniques on difficult problems such as "as optimizations involving discontinuous, noisy, high-dimensional, and multimodal objective functions." Adding, that base Genetic Algorithms their effectiveness on their "ability to exploit efficiently this vast amount of accumulating knowledge by means of relatively simple selection mechanisms".

“The most attractive feature of GA is that it has the ability to explore the search space with the help of the entire population of individuals.” (Hussain and Muhammad, 2020)

Further advantages are the "high implicit parallelism", which makes genetic algorithms "numerically very efficient" (Marsili Libelli and Alba, 2000).

“The power and simplicity of GA make it popular for even largescale optimization problems. The main advantage of GA is that it does not require neither mathematical expression of response surfaces nor any derivative or gradient information” (Boyabatli and Sabuncuoglu, 2004)

Basic Structure At its core, a genetic algorithm uses a list of simple evolution inspired functions. A basic genetic algorithm can be seen here, according to Srinivas and Patnaik, 1994.

Probably ref Holland or Dejong

```
1    simple_genetic_algorithm()
2    {
3        initialize_population();
4        evaluate_population();
5        for(int i = 0; i < num_of_generations; i++)
6        {
7            select_individuals_for_next_population();
```

```
8               perform_crossover();
9               perform_mutation();
10              evaluate_population();
11          }
12      }
```

First, the population is initialized randomly and its individuals are subsequently evaluated using the fitness function. The following steps are now repeated until some stopping criteum is triggered (Grefenstette, 1986) (in this case, it is just a maximum on the number of generations). The individuals are chosen using a selection operation, which takes their fitness value under consideration. Next, under some crossover probabilty, the selected individals mate, producing new offspring. This surfs the purpose of exchanging information between the chromosomes as genes. Futher, a handfull of indivduals (selected using a mutation probabilty) get mutated, which adds small variations to the chromomsomes, thus adding new variation to the population. Finally the new population gets evaluated.

Problems According to Hussain and Muhammad, 2020, genetic algorithms suffer from an exploration vs exploitation dilemma. If the algorithm converges too quickly to a solution, most of the search space might not have yet been explored, thus increasing the probability of getting stuck at only a local optimum. In contrast to that, a low convergence rate will be time consimung and result in wasted compuational resources.

“Finding a balance between exploration and exploitation has been a difficult-to-achieve goal from the beginning.” (K. De Jong, 2007)

Genetic algorithms also require an application specific encoding of genes, which requires time and domain specific knowledge. Due to know real consus on the "best parameter settings" (see section 4, their optimal selection proves to be difficult.

" Premature convergence is a common issue for GA. It can lead to the loss of alleles that makes it difficult to identify a gene [15]. Premature convergence states that the result will be suboptimal if the optimization problem coincides too early. "Katoch, Chauhan, and Kumar, 2021

2.1.1 Genes

The encoding of the genes proved to be, besides the hyperparamter tuning, the main challenge of setting up a genetic algorithm pipeline. A simple and commonly used representation for genes is binary encoding. As suggested by its name, a gene can take the for as 0 or 1. Futher common encodings are octal and hexadecimal representations (Srinivas and Patnaik, 1994, Katoch, Chauhan, and Kumar, 2021).

In case a custom encoding needs to be used, the genetic operation crossover and mutation have to be tailored accordingly. Srinivas and Patnaik, 1994 suggests, to use integer representations in case a optimization problem has real-valued continuous variables. Theses variables are linearly mapped to integers defined in a specific range. It is also possible to again represent these integers as binary encodings.

Various other encodings are available, often very problem specific. For example tree encodings allows for genes to represent programming functions, leading to a path of genetic algorithms called "genetic programming" (Katoch, Chauhan, and Kumar, 2021).

2.1.2 Control Parameter of a Genetic Algorithm

Hyperparamter have a huge influence on the performance of a Genetic Algorihm. They have an impact on the "convergin" ... It has been shown, that there is no universal hyperparamter set and that it needs to be optimized on a per "problem" basis.

Num of generations

The Number of Generation defines the duration of a GA. As long as the algorihtm has not converged, ...? For my testing, using a generation size of 30 was almost always sufficient, and will thus mostly be used.

Population Size

Controlling the Population Size, which defines the number of individuals per generation, has a direct impact on the performance of the genetic algorithm. Research seams to be in agreement, that a small population leads to less diverse individuals and might provide an insufficient sample size. This can lead to a premature convergence to a local optimum. Increasing the population size will allow the genetic algorithm to perform a more informed search. However computation time will suffer due to the larger number of individuals per generation as well as slower convergence to an optimum (Grefenstette, 1986, Katoch, Chauhan, and Kumar, 2021, K. De Jong, 2007). K. De Jong, 2007 also suggest, that "complex, multi-peaked landscapes may require populations of 100s or even 1000s of parents in order to have some reasonable chance of finding globally optimal solutions" This Exploration-Exploitation problem is at the core of the Genetic Algorithm.

Increasing population size will increase the degree if parallelism, as each individual represents one search point (Mills, Filliben, and Haines, 2015).

Selection

The selection operator chooses two parents for crossover and mutation opteriation until the list offsprings has reached the desired population size. The selection operator will favour individuals based on their fitness value, ensuring their increased representation from generation to generation (Srinivas and Patnaik, 1994). Weak solutions thus will be discarded over time. Different selection methods like stochastic uniform remainder, random selection, rank selection, roulette wheel selection and tournament selection exist (Majumdar and Ghosh, 2015). The choice of selection method highly affects the performance of the Genetic Algorithm (Hussain and Muhammad, 2020).

The selection algorithm again needs to satisfy two requirements. On the one hand, high selection pressure will lead to decreased diversity in the population resulting in premature convergence (Katoch, Chauhan, and Kumar, 2021). This results in the algorithm behaviong more like a local search method like a hill-climber or a “greedy” algorithm (K. De Jong, 2007).

The initial low average fitness value of the population will in combination with a few good performing individuals will lead to them overtaking the population, drastically reducing diversity. On the other hand, a low selection pressure will struggle to converge at an optimum.

Hussain and Muhammad, 2020 propose a selection method where initially a low selection pressure is applied, while increasing it to the end, arguing, that "trade-off between these two competing criteria, an adjustable selection pressure must desired".

" The proportionate selection scheme allocates approximately equal numbers of offspring to all strings, thereby depleting the driving force that promotes better strings. Scaling mechanisms and rankbased selection schemes overcome these two problems. "Srinivas and Patnaik, 1994

ROULETTE

" The first selection mechanism for GA was fitness proportional selection (FPS), which was introduced by Holland [1]. Now, it has become the most prevalent selection approach which used the concept of proportionality. It works as the fitness value of each individual in a population corresponds to the area of roulette wheel proportions. Then, an individual is marked by the roulette wheel pointer after it has spun. This operator gives individuals, a probability pi of being selected Eq. (1) that is directly proportionate to their fitness: However, the difficulty is encountered when a significant difference appears in the fitness values [14,17,18]. The scaling problem which is the major drawback of this scheme was first pointed out by Grefenstette [19]. It has happened when population evolves, the ratio between the variance and the fitness average becomes increasingly small. The selection pressure, therefore, drops as the population converges [7]. On the other hand, high selection pressure may lead to premature convergence to a sub-optimal solution. "Hussain and Muhammad, 2020 " Roulette wheel selection maps all the possible strings onto a wheel with a portion of the wheel allocated to them according to their fitness value. This wheel is then rotated randomly to select specific solutions that will participate in formation of the next generation [88]. However, it suffers from many problems such as errors introduced by its stochastic nature. "Katoch, Chauhan, and Kumar, 2021 " De Jong and Brindle modified the roulette wheel selection method to remove errors by introducing the concept of determinism in selection procedure.

Rank selection is the modified form of Roulette wheel selection. It utilizes the ranks instead of fitness value. Ranks are given to them according to their fitness value so that each individual gets a chance of getting selected according to their ranks. Rank selection method reduces the chances of prematurely converging the solution to a local minima [88]. "Katoch, Chauhan, and Kumar, 2021

TOURNAMENT " The tournament selection (TS) is also widely used as an alternative to FPS. In TS, first, randomly select the t (where t is the predefined tournament size) individuals from the population and then they compete against each other based on their fitness. An individual with higher fitness value is declared as a winner and selected for mating process. The selection pressure can be adjusted with change the tournament size [7]. Usually, the most used tournament size is 2 (binary tournament selection (BTS)), which is the simplest form of TS [21]. However, the larger tournament size can be used to enhance the competition among individuals, but it leads to loss of population diversity [22,23]. "Hussain and Muhammad, 2020

" Tournament selection technique was first proposed by Brindle in 1983. The individuals are selected according to their fitness values from a stochastic roulette wheel in pairs. After selection, the individuals with higher fitness value are added to the pool of next generation [88]. "Katoch, Chauhan, and Kumar, 2021

"Qualitative analysis of the selection strategies is depicted, and the numerical experiments show that SGA with tournament selection strategy converges much faster than roulette wheel selection." Jinghui Zhong et al., 2005 "

Roulette wheel selection is the most frequently used selection strategy. "Jinghui Zhong et al., 2005 " Tournament selection is also a selection strategy which selects individuals based on their fitness value. The basic idea of this strategy is to select the individual with the highest fitness value from a certain number of individuals in the population into the next generation. In the tournament selection, there is no arithmetical computation based on the fitness value, but only comparison between individuals by fitness value. The number of the individuals taking part in the tournament is called tournament size. "Jinghui Zhong et al., 2005

" 1) Randomly select several individuals from the population to take part in the tournament. Choose the individual that has the highest fitness value from the individuals selected above by comparing the fitness value of each individual. Then the chosen one is copied into the next generation of the population. 2) Repeat step1 n times where n is the number of individuals of the population. "Jinghui Zhong et al., 2005 " From the results shown above, SGA using tournament selection always obtains the satisfied solutions with more times at earlier generations than roulette wheel selection. This indicates SGA based on tournament selection converges more quickly than roulette wheel selection. "Jinghui Zhong et al., 2005

OTHER " The most popular technique is the linear rank selection (LRS) scheme proposed by Baker [20]. It sorts the individuals in the sequence as worst to best according to the fitness and allocates them a survival probability proportional to their rank order. After this task, a sampling procedure (i.e., roulette wheel sampling) is used to select the individuals for mating process. In this way, the LRS can maintain a constant selection pressure throughout in the sampling process, because it introduces a uniform scaling across the population. The weakness of this scheme is that it can lead to slower convergence, because there is no significant difference between the best and other individuals. The selection probability of two consecutive chromosomes by the same amount is regardless of whether the gap between their fitness is larger or smaller [7]. "Hussain and Muhammad, 2020 "

Another rank-based selection scheme is exponential ranking selection (ERS). It works similar as to LRS, except for the non-linear assignment of probabilities to the individuals. "Hussain and Muhammad, 2020

" The main contribution of this article is in the development of the proposed selection approach which reduces the weakness associated with FPS and LRS in the GA procedure. The proposed approach is based on the ranking scheme which splits the individuals after ranking and then assign them probabilities for selection. This will increase the competition among individuals to be selected for mating process to regulate the selection pressure. "Hussain and Muhammad, 2020

" The LRS introduces slow convergence speed and sometimes converges to a sub-optimal solution as less fit individuals may be preserved from one generation to another. In GA, the FPS has the essence of exploitation, while LRS

is influenced by exploration. The information about the relative evaluation of individuals is ignored, all cases are treated uniformly regardless of the magnitude of the problem and, finally, the schema theorem is violated. LRS prevents too quick convergence and differs from FPS in terms of selection pressure. This discussion suggests that, whenever a selection procedure is used, some kind of adaptation of the selection pressure is highly desirable. "Hussain and Muhammad, 2020

" In this research, we propose an alternative selection scheme [split rank selection (SRS)] that maintains a fine balance between exploration and exploitation.

In the proposed procedure, the individuals are ranked according to their fitness scores from worst to best, thus overcoming the fitness scaling issue. After this, split the whole population into two portions and assigning them probabilities for selection based on their ranks. "Hussain and Muhammad, 2020

" Stochastic universal sampling (SUS) is an extension to the existing roulette wheel selection method. It uses a random starting point in the list of individuals from a generation and selects the new individual at evenly spaced intervals [3]. It gives equal chance to all the individuals in getting selected for participating in crossover for the next generation. Although in case of Travelling Salesman Problem, SUS performs well but as the problem size increases, the traditional Roulette wheel selection performs relatively well [180]. "Katoch, Chauhan, and Kumar, 2021

" Boltzmann selection is based on entropy and sampling methods, which are used in Monte Carlo Simulation. It helps in solving the problem of premature convergence [118]. "Katoch, Chauhan, and Kumar, 2021

" An alternate way to avoid the twin problems that plague proportional selection is rank-based selection, which uses a fitness value-based rank of strings to allocate offspring. The scaled fitness values typicallyvary linearly with the rank of the string. The absolute fitness value of the stringdoes not directlycontrol the number of its offspring. To associate each string with a uniquerank,this approach sortsthe stringsaccordingto their fitnessvalues,introducing the drawback of additional overhead in the GA computation. Another mechanism is tournament selection. For selection,a stringmust win

a competition with a randomly selected set of strings. In a k-arytournament, the best of k stringsis selectedfor the next generation. "Srinivas and Patnaik, 1994

ELITE " However, there is a possibility of information loss. It can be managed through elitism [175]. Elitism selection was proposed by K. D. Jong (1975) for improving the performance of Roulette wheel selection. It ensures the elitist individual in a generation is always propagated to the next generation. If the individual having the highest fitness value is not present in the next generation after normal selection procedure, then the elitist one is also included in the next generation automatically [88]. "Katoch, Chauhan, and Kumar, 2021

Crossover

The crossover operation serves as the mating process between two individuals (chosen from the population by the selection operation). Its resulting offspring has a combination of both parents genes. Grefenstette, 1986 higlights, that this "First, it provides new points for further testing within the hyperplanes already represented in the population." and "Second, crossover intro- duces representatives of new hyperplanes into the population".

Different types of crossovers can be chosen, the most simple approach being the single-point crossover. Here a randomly chosen point aloong the length of the chromosome is chosen. The two parents will swap their genetic information at this point, generating two new offsprings (Katoch, Chauhan, and Kumar, 2021). Extenations are the two- or multipoint crossover operations, where chromosomes are divided by k segments, which will get exchanged. They eliminate a disadvantage of "the single-point crossover bias toward bits at the ends of strings." (Srinivas and Patnaik, 1994)

A second crossover opteration is the uniform crossover, here, based on a probability, it randomly decides to swap genes between the parents, independent of the exchange of other genes (Katoch, Chauhan, and Kumar, 2021).

enquoteTo classify techniques, we can use the notions of positional and distributional biases. A crossover operator has positional bias if the probability

that a bit is swapped depends on its position in the string. Distributional bias is related to the number of bits exchanged by the crossover operator. If the distribution of the number is nonuniform,the crossover operator has a distributional bias. Among the various crossover operators, single-point crossover exhibitsthe maximum positional bias and the least distributional bias. Uniform crossover, at the other end of the spectrum, has maximal distributional bias and minimal positional bias. (Srinivas and Patnaik, 1994)

According to Srinivas and Patnaik, 1994 the ignoring of the genes position results in a higher disruptiveness, which has potential drawbacks. However it will be more exporatory in homogeneous populations where k-point crossovers might struggle to explore. Srinivas and Patnaik, 1994 also suggests that uniform crossover is more usefull in small populations. Larger populations inherently are more diverse, makeing k-point crossover more suitable.

Other crossover operations are analyzed in paper .

Find good research paper, maybe Katoch, Chauhan, and Kumar, 2021

Not only the crossover type has to be chosen, a crossover probility (or crossover rate) also needs definition. This value defines, how likely a single crossover operation is to be applied on the parent population. If no crossover will be applied, the parents will directly be used as children. The problem of choosing a crossover probability again comes back to the exploration/-exploatation balance. Higher crossover rates will introduce new structures quickly into the population, however good structures might get disrupted. A low crossover rate will result in a low exploration, resulting in stagnation (Grefenstette, 1986).

“If reproductive variation is too strong, the result is undirected random search.” (K. De Jong, 2007)

Mutation

Mutation is responsible for introducing new information into the gene pool.

" Setting the mutation probability too high may transform the search procedure into random testing, but it also helps to introduce new gene material,

which promotes the exploration of new input space regions. "Klampfl, Klück, and Wotawa, 2023

" Mutation is a secondary search operator which increases the variability of the population. "Grefenstette, 1986

" A low level of mutation serves to prevent any given bit position from remaining forever converged to a single value in- the entire population. A high level of mutation yields an essentially random search. "Grefenstette, 1986

" With a larger population and higher mutation rate, the population will tend to contain more variety, thus increasing the random aspects of the GA. "Grefenstette, 1986

" The absence of mutation is also associated with poorer performance, which suggests that mutation performs an important service in refreshing lost values. "Grefenstette, 1986

" After crossover. strings are subjected to mutation. Mutation of a bit involves flipping it: changing a 0 to 1 or vice versa. Just asp, controls the probability of a crossover, another parameter. P,, (the mutation rate), gives the probability that a bit will be flipped.The bits of a string are independently mutated that is, the mutation of a bit does not affect the probability of mutation of other bits. The SGA treats mutation only as a secondary operator with the role of restoring lost genetic material. "Srinivas and Patnaik, 1994

" Increasing the mutation probability tends to transform the geneticsearch into a random search, but it also helps reintroduce lost genetic material. "Srinivas and Patnaik, 1994

" Mutation is not a conservative operator and can generate radically new buildingblocks. "Srinivas and Patnaik, 1994

" Mutation: After crossover, the springs are subjected to mutation. Mutation functions make small random changes in the individuals in the population, which provide genetic diversity and enable the genetic algorithm to search a broader space. The different forms of mutation are constraint dependent, uniform, adaptive feasible etc. Mutation of a bit involves flipping it, changing between 0 to 1 and vice versa with a small mutation probability. "Majumdar and Ghosh, 2015

" Mutation is an operator that maintains the genetic diversity from one population to the next population. "Katoch, Chauhan, and Kumar, 2021

" If the mutation is not considered during evolution, then there will be no new information available for evolution. "Katoch, Chauhan, and Kumar, 2021

"In Genetic Algorithms mutation probability is usually assigned a constant value, therefore all chromosome have the same likelihood of mutation irrespective of their ®tness."Marsili Libelli and Alba, 2000

" Mutation introduces random binary changes in a chromosome. Usually the mutation likelihood is kept constant at a low value for all bits. In this paper, a new mechanism of mutation will be introduced and the consequences of this modi®cation assessed. "Marsili Libelli and Alba, 2000

Adaptive Mutation "It is shown in this paper that making mutation a function of ®tness produces a more ef®cient search. This function is such that the least signi®cant bits are more likely to be mutated in high-®tness chromosomes, thus improving their accuracy, whereas low-®tness chromosomes have an increased probability of mutation, enhancing their role in the search. In this way, the chance of disrupting a high-®tness chromosome is decreased and the exploratory role of low-®tness chromosomes is best exploited. " Marsili Libelli and Alba, 2000

" The weak point of "classical" GAs is the total randomness of mutation, which is applied equally to all chromosomes, irrespective of their fitness. Thus a very good chromosome is equally likely to be disrupted by mutation as a bad one. On the other hand, bad chromosomes are less likely to produce good ones through crossover, because of their lack of building blocks, until they remain unchanged. They would benefit the most from mutation and could be used to spread throughout the parameter space to increase the search thoroughness. So there are two conflicting needs in determining the best probability of mutation. Usually, a reasonable compromise in the case of a constant mutation is to keep the probability low to avoid disruption of good chromosomes, but this would prevent a high mutation rate of low-fitness chromosomes. Thus a constant probability of mutation would

probably miss both goals and result in a slow improvement of the population. "Marsili Libelli and Alba, 2000

" This paper has presented an improved search algorithm to reduce the chance that high-®tness chromosomes are muted during the search, thus losing their favourable schemata. In GAs, mutation is usually assigned a constant probability and thus all chromosome have the same likelihood of mutation irrespective of their ®tness. Conversely, making mutation a function of ®tness produces a more ef®cient search. "Marsili Libelli and Alba, 2000

" signi®cant bits are likely to be muted in high-®tness chromosomes, thus improving their accuracy, whereas low-®tness chromosomes are much more likely to change, enhancing their exploratory role in the search. "Marsili Libelli and Alba, 2000

" In all cases the new algorithm showed a faster improvement of the maximum ®tness, which was always greater than the one produced by the classical GA at the same generation. "Marsili Libelli and Alba, 2000

" In 2007, DeJong took a broader view, considering what was known with respect to the wider community of evolutionary algorithms (EAs). He observed that while it appears that adapting mutation rate on-line during execution provides advantages, most EAs are deployed with a default set of static parameter values that have been found quite robust in practice. "Mills, Filliben, and Haines, 2015

" Parameter settings optimal in the earlier stages of the search typically become inefficient during the later stages. Similarly, encodings become too coarse as the search progresses, and the fraction of the search space that the GA focuses its search on becomes progressivelysmaller. To overcome these drawbacks, several dynamic and adaptive strategiesfor varying the control parameters and encodings have been proposed. One strategy exponentially decreases mutation rates with increasing numbers of generations, to gradually decrease the search rate and disruption of strings as the population converges in the search space. Another approach considers dynamically modifying the rates at whichthe various genetic operators are used, based on their performance. Each operator is evaluated for the fitness values of strings it generates in subsequent generations. Very often, after a large fraction

of the population has converged (the strings have become homogeneous), crossover becomes ineffective in searching for better strings. Typically, low mutation rates (0.001 to 0.01) are inadequate for continuing exploration. In such a situation, a dynamic approach for varying mutation rates based on the Hamming distance between strings to be crossed can be useful. The mutation rate increasesas the Hamming distancebetween strings decreases. As the strings to be crossed resemble each other to a greater extent,the capacity of crossover to generate new strings decreases, but the increased mutation rate sustainsthe search. "Srinivas and Patnaik, 1994

" The purist might argue that inventing feedback control procedures for EAs is a good example of over-engineering an already sophisticated adaptive system.

A better strategy is to take our inspiration from nature and design our EAs to be self-regulating. For example, individuals in the population might contain "regulatory genes" that control mutation and recombination mechanisms, and these regulatory genes would be subject to the same evolutionary processes as the rest of the genome. "K. De Jong, 2007

" Perhaps the most interesting thing to note today, after more than 30 years of experimenting with dynamic parameter setting strategies, is that, with one exception, none of them are used routinely in every day practice. The one exception are the strategies used by the ES community for mutation step size adaptation. "K. De Jong, 2007

" My own view is that there is not much to be gained in dynamically adapting EA parameter settings when solving static optimization problems. The real payoff for dynamic parameter setting strategies is when the fitness land- scapes are themselves dynamic (see, for example, [4], [5], or [32]). "K. De Jong, 2007

" For instance, in a GA, one may want to decrease the probability of crossover and mutation over time to avoid too high reproductive variation hindering convergence to local optima. However, in that case, one would need to set new parameters that define how much and how often each probability is decreased. Research has not yet concluded if one technique has a clear advantage over the other. "Klampfl, Klück, and Wotawa, 2023

Other

Diversity " In initial stage of GA, the similarity between individuals is very low. The value of R should be low to ensure that the new population will not destroy the excellent genetic schema of individuals. At the end of evolution, the similarity between individuals is very high as well as the value of R should be high. "Katoch, Chauhan, and Kumar, 2021

Fitness Function " The Fitness Function: The fitness function in Genetic Algorithm represents the objective function and the fitness value corresponds the performance of an individual chromosome. "Majumdar and Ghosh, 2015

" Multiobjective GA (MOGA) is the modified version of simple GA. MOGA differ from GA in terms of fitness function assignment. The remaining steps are similar to GA. The main motive of multiobjective GA is to generate the optimal Pareto Front in the objective space in such a way that no further enhancement in any fitness function without disturbing the other fitness functions [123]. "Katoch, Chauhan, and Kumar, 2021

" The concept of Pareto dominance was introduced in multiobjective GAs. Fonseca and Fleming [56] developed first multiobjective GA (MOGA). The niche and decision maker concepts were proposed to tackle the multimodal problems. However, MOGA suffers from parameter tuning problem and degree of selection pressure. "Katoch, Chauhan, and Kumar, 2021

" The majority of applications of optimization tools to subsurface remediation problems have been based on single objective optimization methods. Single objective methods can accommodate multiobjective problems in several ways, such as minimizing a weighted, linear combination of the objective functions or minimizing a single objective while transforming the remaining ob- jectives into constraints. However, these methods rely on a priori knowledge of the appropriate weights or constraint values. Furhtermore they are only capable of findig individual points on the tradeoff curve(or sureace) for each problem solution.

As already mentioned, previous approaches for solving the multiobjective problem have involved re- ducing the problem dimension, either by combining all objectives into a single objective (e.g. [27]) or optimizing one while the rest are constrained (e.g., [3]). "Erickson, Mayer, and Horn, 2002

" True multiobjective methods have the potential to simultaneously generate all possible optimal combina- tions of objectives, with less effort than other ap- proaches. Multiobjective problems involve several objective functions, each of which is a function of de- cision ðdÞ and state variables ðsÞ. tiveproblemcanbestatedas: "Erickson, Mayer, and Horn, 2002

" Multiobjective approaches in this category operate on the concept of "Pareto domination", which states that one candidate dominates another only if it is at least equal in all objectives and superior in at least one "Erickson, Mayer, and Horn, 2002

More on multiobjecte approaches can be found in Erickson, Mayer, and Horn, 2002.

Stopping Criteria " Stopping criteria: Stopping criteria determines what causes the algorithm to terminate-generations, time limit, fitness limit etc. "Majumdar and Ghosh, 2015

This is not used for this GA

2.2 Behavior Tree

A behavior tree is a decision tree.

insert a good introduction to BT

2.2.1 Usage for GA

Due to the fact, , that there is no full stack available for the EGO vehicle, a solution had to be found. In order to have the Genetic Algorithm controll only NPCs and not the EGO vehicle itselve, a behaviour tree is used. The behaviour tree is used to controll the EGO vehicle over the action interface

insert ref to discussion

provided by the Traffic Manager. This is the same as the Genetic Algorithm is doing.

The behaviour tree will define which direction the EGO should take at junctions and it will realistically dodge obstacles intoduced by the Genetic Algorithm. The main goal of the BT is to make the EGO vehicle behave in a realistic way.

In a further chapter it will be dicussed if a GA with controll of the EGO (i.e. no BT will be used) lead to better cost.

While the aim of the GA is to find the most optimal solution, considering the vastness of the hyperspace, this is unlikely. Rather, we want to find the "best" local minimas. Considering the contex of Automotive testing, it is not so much of importance to find "the best fail of the ADAS/AD System", rather its important to find "all" fails.

3 Implementation

3.1 Traffic Manager

The Genetic Algorithm will control the simulation of a custom developed Traffic Manager. This Traffic Manager was developed closely to fit the needs of the Genetic Algorithm. It, however is not part of this Thesis and will thus will only get a brief introduction. In general, it will simulate traffic starting from a predefined scenarios which defines the positions and types of vehicles and pedestrians(i.e. actors). A simulation always consist of at least one EGO vehicle. Additionally any number of NPCs can be used.

While the NPCs are only controlled by the Traffic Manager, the ego vehicle can be either partly or even completely controlled by an ADAS/AD Function. The stated goal is to test these functions for errors.

For all simulations done by this thesis, the Traffic Manager was set to 100Hz.

3.1.1 Action Interface

To control the behaviour of the actors inside the simulation, actions can be requested over the "Action Interface" provided by the Traffic Manager. An action will request a certain behaviour from an actor. An action can be set to at any timestep for any actor[1]. Pedestrians and vehicles have a different set of actions.

Insert graph of action interface

[1]depending on the ADAS/AD function under test, the Action Interface might be disabled for the EGO vehicle

If no action is set, the actor will behave in a normal manner inside the simulation. This means that the actor will follow along its path until a new action changes its behaviour.

The following list are all actions provided by the traffic manager that were available for the genetic algorithm at the time of this master thesis.

- JunctionSelection
  - Parameters: Vehicle ID: int, Junction_selection_angle: float
  - Angle is set in radiant. Default value is 0. Vehicles will chose which direction to take at a junction based on this angle.
- LaneChange
  - Parameters: Vehicle ID: int, ...
  - Initiates a LaneChange based on its given parameters.
- AbortLaneChange
  - Parameters: Vehicle ID: int, ...
  - If a LaneChange is currently happening, it will get aborted.
- ModifyTargetVelocity
  - Parameters: Vehicle ID: int, ...
  - Modifies the interal Target Velocity of the Traffic Manager by a percentage. If it is for example 0, the vehicle will stop.
- TurnHeading
  - Parameters: Pedestrian ID: int, ...
  - The pedestrian will turn 180 degrees and walk in the oposite direction
- CrossRoad
  - Parameters: Pedestrian ID: int, ...
  - The pedestrian will cross the road immediately.
- CrossAtCrosswalk
  - Parameters: Pedestrian ID: int, ...
  - The pedestrian will cross the road at the next crosswalk.

All these actions are accessed by the Genetic Algorithm and the Behavior tree. The Behaviour tree sets only actions for the EGO vehicle, while the Genetic Algorithm will set all actions for the other actors in the simulation.

3.2 Genetic Algorithm

The task of the Genetic Algorithm is to search for sequences of actions that will result in the most interesting Scenarios according to its cost function. For implementing the Genetic Algorithm, DEAP was chosen. It is a popular tool for academia and allows for high customacibility. As has been stated in section 3.1.1, it has full access over setting actions for all NPCs. Using theses actions, it tries to optimize a cost function. This section aims to explain how the genetic algorithm was implemented and which different hyperparameters are variable. In chapter 4, the best hyperparameter combination will be generated, further analysis is done in 5.

cite

cite 3 examples

A few default settings for the genetic algorithm had to be chosen. It was decided that the genetic algorithm will set an action per actor every 50 steps, which translates to 0.5 seconds (simulation runs at 100hz). In other words, every 50 steps of the simulation is 1 timestep for the genetic algorithm. The time of a simulation is always set to 35 seconds. Each genetic algorithm will run for 30 generations. These to settings were chosen with the aim of reducing the amount of needed computations.

If the GA decides to not set an action for the integer, it sets "NoAction" as a placeholder.

3.2.1 Encoding

When implementing a Genetic Algorithm, it is necessary to implement an encoding that fits to the problem. Each individual basically thus needs to include all actions that the genetic algorithm wants to apply. Different encodings presented in section 2.1.1, however none directly fitted to the problem presented. A custom encoding for both chromosomes and genes needed to be generated.

Chromosome

ref

Each individual has 1 chomosome which consits of a list of genes, which has been explained in section . Starting out, 2 different encodings came to mind, in both cases, the genes position in the chromosome defined the time an action is set.

Time The first encoding is will be called "Time". Each gene corresponds to 1 timestep (so 1 gene per every 0.5 seconds). One gene has a list of the length of the number of all NPCs. This list is populated with actions. The index of an action in the list corresponds to the NPC id (index + 1 as the ego has id == 0, thus a start at 1 is needed). A visualization is seen in figure 3.1.

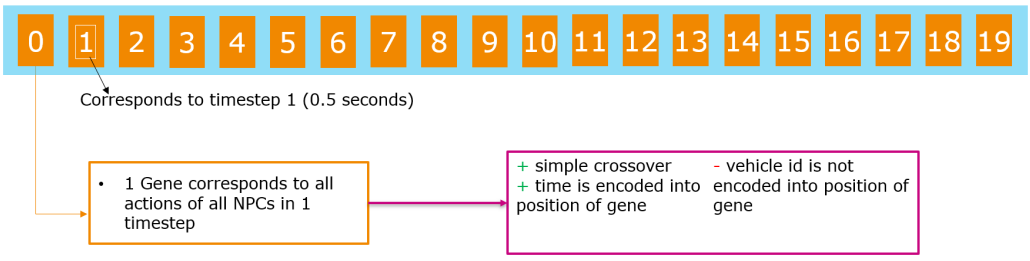

Figure 3.1: Time

Given the previously stated simulation time of 35 seconds, each chromosome has a length of $35 * 2 = 70$ genes. Each gene consits of $number\_of\_actors$ actions. Crossover can thus only move all actions of a timestep at once, modifing between actions of the same timestep can only be done using mutation. If this is desired will be seen in the next chapters.

TimeNPC The second encoding has the name "TimeNPC", and is somewhat differently structured. Now, genes only hold 1 action, encoding now not only the timestep, but also the actor id in the position of the gene inside

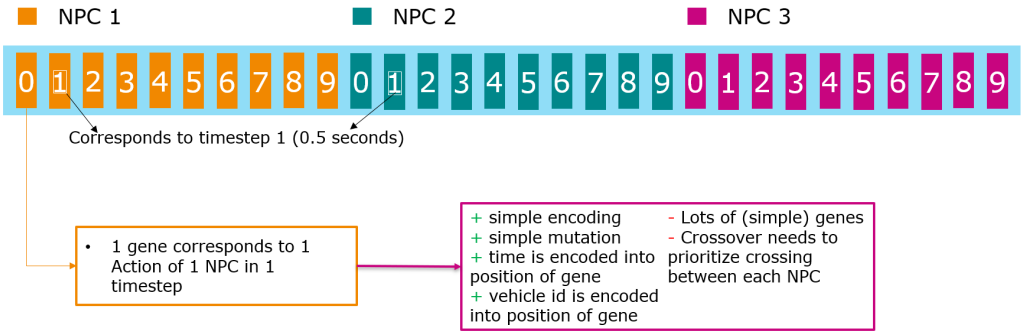

Figure 3.2: Time + NPC

the chromosome. Now, each actors actions will be listed one after another. This is visualized in figure 3.2.

Now, each gene has a length of 1 and each chromosome now has a length of $35 * 2 * number\_of\_actors$, which makes them much longer compared to the previous encoding. This now allows the crossover operation to modify only specific actions of one timesstep. Previously this was not possible.

However for this encoding to make sense, the crossover operations "One-Point" and "TwoPoint" had to be modified as follows. In an example of 10 NPCs, the operations will be executed for each NPC separately. Otherwise these two operations would have only had an effect on 1 or 2 different NPCs. For the reaiming NPCs, their actions would stay the same.

Gene

Two different encodings for genes were implemented as well. A gene always consits of a list, which depending on the chromosome type either has a length of $number\_of\_actors$ (In case ChromosomeEncoding == Time) or of length 1 (in case ChromosomeEncoding == TimeNPC). The following two encodings thus show the type of object, which is in these lists.

Integer The first encoding uses integer, which are translated into actions when the simulation is started. For each action, a range of integers is assigned, the larger the range, the more likely the action is chosen by the GA. Actions that have parameters are split into different ranges, according to which paramters make sense. For example ModifyTargetVelocity is split into five different parts, with different percentages, namely 50, 70, 100, 130, 160. The range of integers assigned to these parts is different. A percentage setting of 100 for example has the largest integer range assigned. In Appendix , the probabilty of an actions can be seen. In Appendix ... the probabitly per actions of the parameters can be viewed.

These ranges were assigned based on intuition and trial and error. The encoding is visualized in 3.3.

TODO: Due to the cliff problem, it was decided to not use binary representations for these integers. Due to the mutation choosing the integers randomly, it is not a problem that the variables are not really "continous".

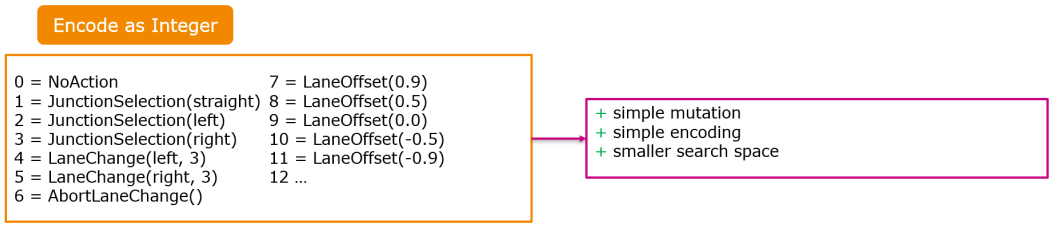

Figure 3.3: Integer

Dictionary The second encoding is much similar to the actual actions used in the simulations. Now, no translation is necessary anymore. During generation of the individuals, each action is again selected based on different probabilities assigned to actions, which again can be viewed in Appendix ... . These probablilites are the same as for the integer encoding. However the difference is, in case an action has parameters that need to be chosen. For each parameter, a range and a randomness function was chosen. For example in case of the percenetage parameter in ModifyTargetVelocity, the values are selected from a GausDistribution, with mu= 100, sigma=25 and a range limit between 0 and 300.

Again, these probability functions with settings were assigned based on intuition as well as trial and error. Detailed information can be seen in Appendix...

Figure 3.4 shows a visualization.

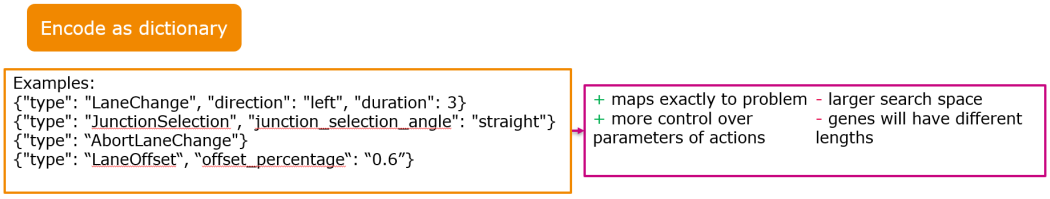

Figure 3.4: Dictionary

3.2.2 Cost Function

Suggestion florian: better explain cost values (e.g. 3500 means no emergency break...)

Cost function is a bit difficult, as we are only using internal values. No ADAS/AD system is tested and we thus have to work with what we got. This is the code of the cost function:

```
SEPS_PER_SECOND = 100
# allow emergency breaks to last only 3 seconds
MAX_DURATION = 3 * STEPS_PER_SECOND

cost = 0
duration_counter = 0
for i in range(len(result["ego_emergency_stop"])):
    if not result["ego_emergency_stop"][i]:
        # base cost for no current emergency break
        cost = cost + 1
        duration_counter = 0
    else:
        if duration_counter > MAX_DURATION:
            # increase cost if emergency break max
               duration is exceeded
            cost = cost + 10
        duration_counter += 1
return cost
```

result["ego_emergency_stop"] is a list with the length $100 * simulation_duration_seconds$ (because 100hz). It contains a boolean per step, if the EGO vehicle has initiated an emergency stop.

It would have been interesting to not only test for emergency stops (which will make the NPCs try to get the EGO to hard break often) but also improve time to collison (TTC), as was done by . However by the time of starting the testing, no working TTC functionality was implemented. Thus, only the emergency break cost function is used by the GA to be optimized.

Ref florian

3.3 Behavior Tree

Depending on the functionality under test, it is possible to let the EGO vehicle be controlled by a Behaviour Tree. This makes sense if for example a functionilty like AEB is tested, where only the breaks are controlled. In case of a full driving stack, no Behaviour Tree would be used.

The general idea is to have an EGO vehicle moving in a "relateable" manner trough the world. It will try to dodge standing or slow moving obstacles. This needs to be done in a determinisitc manner in order to no introduce randomness into the simulation.

For this, Behaviour Tree is used. While it has access to the same Action Interface (described in section 3.1.1) as the Genetic Algorithm, it is more tightly integrated with the Traffic Manger. While the Genetic Algorithm only ingests the results generated by the Simulation with the cost function, the Behaviour Tree needs access to internal functions during the simulation. The following figure shows the behaviour tree implemented.

Explain BT

Starting out,

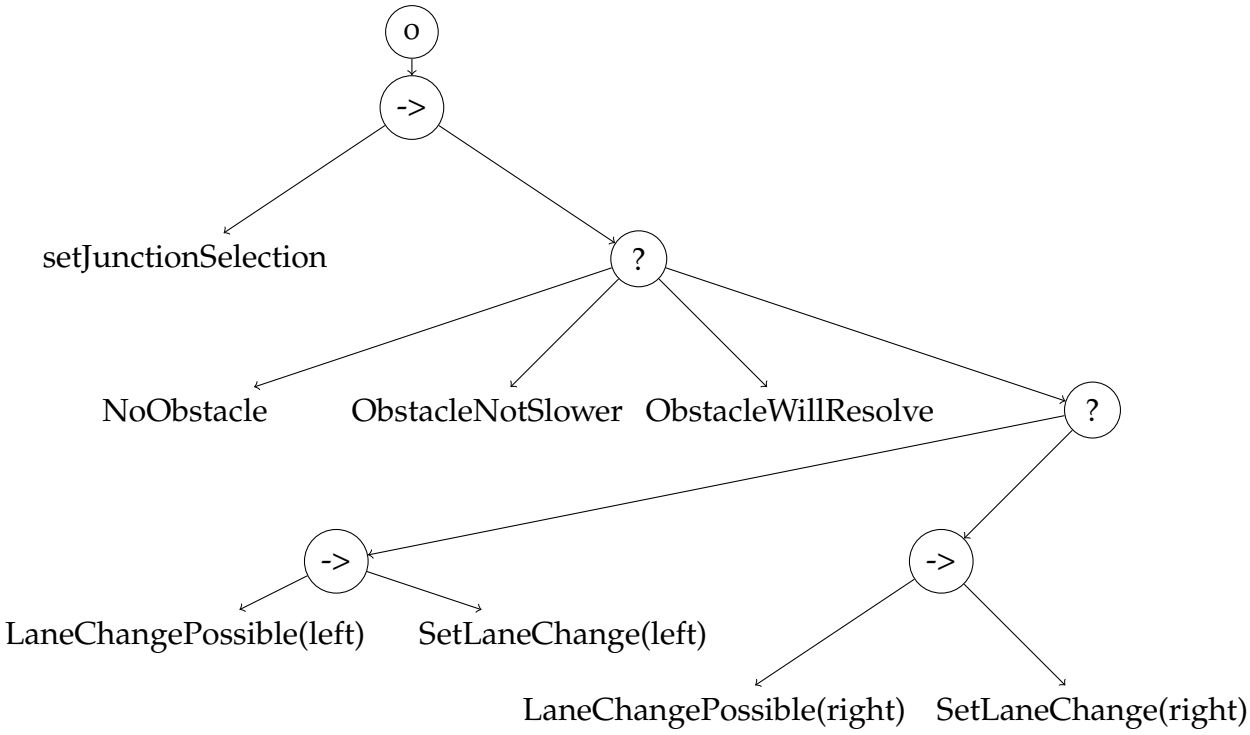

Figure 3.5: Used Behaviour Tree

4 Hyperparameter Tuning

In this chapter, we will incrementally move to an optimized Genetic Algorithm

" changing the population size, the type and amount of reproductive variation, etc. could significantly change the behavior of an EA on a particular fitness landscape and, conversely, appropriate parameter settings for one fitness landscape might be inappropriate for others. "K. De Jong, 2007

4.1 No Free Lunch Theorem

No Free Lunch Theorem: The best hyperparameter settings of a Genetic Algorithm are very problem specific. K. De Jong, 2007, Dao, Abhary, and Marian, 2016

More ref

" It has been shown in a variety of contexts including the areas of search, optimization, and machine learning that, unless adequate restrictions are placed on the the class of problems one is attempting to solve, there is no single algorithm that will outperform all other algorithms (see, for example, [83]).

A more constructive interpretation of NFL results is that they present us with a challenge to define more carefully the classes of problems we are trying to solve and the algorithms we are developing to solve them.

The No Free Lunch results place an obligation on the EA practitioner to understand something about the particular properties of the problems that (s)he is trying to solve that relate to particular choices of EA parameter settings. This poses a bit of a dilemma in that often an EA is being used precisely because of a lack of knowledge about the fitness landscapes

being explored. As a consequence, the parameter setting strategy frequently adopted is a multiple run, human-in-the-loop approach in which various parameter settings are tried in an attempt to fine-tune an EA to a particular problem. "K. De Jong, 2007

" While there is empirical data available to support the choice of appropriate default parameters for a specific problem to be solved,24 it has been formally proven in the “No Free Lunch” (NFL) theorem that there is not one optimal parameter setting for all possible search problems.25 "Klampfl, Klück, and Wotawa, 2023

4.2 Map and Starting Scenario

The map is Town10 from Carla. It was chosen, because 1. its roads are self contained, 2. its not too big, yet still complex and 3. its supported by Carla and thus visualization looks better.

The Starting Scenario defines the number and type of all actors as well as their position. It needs to be created manually. Changing the scenario will have a great impact on the Genetic Algorithms performance. For time and complexity reasons, it was thus decided to first stick with one scenario and do all hyperparameter testing there. And finally test the performance for a handfull different scenarios.

4.3 Population

The number of Individuals is of high importance to a genetic algorithm, as has been explained in section 2.1. Especially considering the limed processing resources available, a suitable population size has to be found. On one hand, a population that is too low might result in less diverse runs of the genetic algorithm, on the other hand, if population is too high, the simulations will become too costly. Considering these points, the first step of the hyper parameter tuning was to find a suitable population size. In the

next chapter 4.4, we will aim to improve the hyperparamter using a more robust approach.

In order to test for the best population size, the other hyperparameters have to be assumed using an educated guess. While reviewing the literature, trends of general settings for genetic algorithms can be found. However Mills, Filliben, and Haines, 2015 highlight the inconsistencies between findings, stating to have "uncovered conflicting opinions and evidence regarding key GA control parameters".

However Grefenstette, 1986 suggests, that "while it is possible to optimize GA control parameters, very good performance can be obtained with a range of GA control parameter settings." This is also complimented by findings from K. De Jong, 2007: "The key insight from such studies is the robustness of EAs with respect to their parameter settings. Getting “in the ball park” is generally sufficient for good EA performance. Stated another way, the EA parameter “sweet spot” is reasonably large and easy to find [18]. As a consequence most EAs today come with a default set of static parameter values that have been found to be quite robust in practice."

Chosing the right selection method is complicated as well, as discuees by K. De Jong, 2007: "One source of difficulty here is that selection pressure is not as easy to “parameterize” as population size. We have a number of families of selection procedures (e.g, tournament selection, truncation selection, fitness-proportional selection, etc.) to choose from and a considerable body of literature analyzing their differences (see, for example, [19] or [15]), but deciding which family to choose or even which member of a parameterized family is still quite difficult, particularly because of the interacting effects with population size [13]."

Looking at the literature might lead to hyperparameters are used that at least sufficient enough, to get an idea which range for population size is suitable. We will now look at different concrete hyperparameter suggestions from the literature.

" At the beginning of an EA run it is important to have sufficient parallelism to handle possible multi-modalities, while at the end of a run an EA is likely to have converged to a local area of the fitness landscape that is no more complex than Figure 1.

As a consequence, most EAs used in practice today run with a fixed population size, the value of which may be based on existing off-line studies (e.g., [34]) or more likely via manual tuning over multiple runs. "K. De Jong, 2007

4.3.1 Suggested hyperparameter from the literature

Use best values also from : Using genetic algorithms for automating automated lane-keeping system testing

In an often cited thesis by K. A. De Jong, 1975, the following parameters have been suggested: GA(50, 0.6, 0.001, 1.0, 7, E) These suggested parameters have been used successfully by various different genetic algorithms Grefenstette, 1986.

Talk about rules (e.g. 1/n for mut rate...) - look at: Parameter selection in genetic algorithms

An extensive study by Mills, Filliben, and Haines, 2015 which that took over "over 60 numerical optimization problems." into consideration found that "the most effective level settings found for each factor: population size = 200, selection method = SUS, elite selection percentage = 8%, reboot proportion = 0.4, number of crossover points = 3, mutation rate = adaptive and precision scaling = 1/2 as fine as specified by the user."

Grefenstette, 1986 claim that GA(30, 0.95, 0.01, 1.0, 1, E) and GA(80, 0.45, 0.01, 0.9, 1, P) produced the best results. They also advised against, a mutation rate of over 0.05, suggesting poor performance. Using a low mutation rate is also suggested by Whitley, 1994 and Jinghui Zhong et al., 2005. On the other hand, Boyabatli and Sabuncuoglu, 2004 state, that "Controversial to existing literature on GA, our computational results reveal that in the case of a dominant set of decision variable the crossover operator does not have a significant impact on the performance measures, whereas high mutation rates are more suitable for GA applications." Other paper also find a relatively high mutation rate useful. Almanee et al., 2021 uses genetic algorithms in a similar domain as this thesis. There, a Population of 50, crossover of 0.8 and mut of 0.2 was used. These used params are the same as the default params from deap (pop = 50 CXPB, MUTPB, NGEN = 0.5, 0.2, 4).

cite https://deap.readthedocs.io/en/master/overview.html

Srinivas and Patnaik, 1994 state, that for a higher population, cross : 0.6, mut: 0.001 and pop: 100 is a good starting point, while a lower population

needs higher crossover and mutation rates like this cross: 0.9, mut: 0.01, pop: 30

Fazal et al., 2005 recommends a population size of 50, a scattered crossover function with a crossover probability of 0.5. The used selection function was tournament selection. Elite count was set to 5.

Dao, Abhary, and Marian, 2016 suggests a population size of 200, two point crossover with a crossover probability of 0.7. A Gausian Mutation Function as well as roulette selection and elite count set to 1.

A population size of 200 and roulette selection is used by Assistant Professor, Amity University, Jaipur, Rajasthan, India et al., 2019. Further, the elite count is set to10. A heuristic crossover function with a crossover probability of 0.4 is also used.

TODO: Jinghui Zhong et al., 2005 set the range as "We set the value of POPSIZE as 50, 100, 150, 200, 250, the value of PXOVER as 0.1, 0.3, 0.5, 0.7, 0.9, and the value of PMUTATION as 0.05, 0.1, 0.15, 0.2, 0.25."

" DeJong believes that EAs pre-tuned with default parameter values for particular problem classes will continue providing better performance than EAs that attempt to dynamically adapt too many control parameters for specific problems. So, even after 30+ years of research, the question of best settings for EA and GA control parameters has no widely agreed answer. "Mills, Filliben, and Haines, 2015

4.3.2 results

This now leads to a difficult decision in choosing the right parameters. Based on the extensive research, we will compare population size of 32, 48, 64 and 96. We will compare the different crossover rates: 0.8 and 0.6. For mutation, 0.01 and 0.2 will be discussed. Further we will use tournament selection with 2 and 4. Each run will be executed 5 times to get rid of randomness and to make the results more robuts. We will run each simulation for 40 Generations.

Comparison of Population Size - mean					
Settings	Code	32	48	64	96
C: 0.6, M: 0.01, TS: 2	A	3051	3016	2851	2871
C: 0.6, M: 0.01, TS: 4	B	3111	3021	3079	2937
C: 0.6, M: 0.2, TS: 2	C	3062	3010	3002	2831
C: 0.6, M: 0.2, TS: 4	D	3020	2967	2891	2850
C: 0.8, M: 0.01, TS: 2	E	3063	2892	2971	2916
C: 0.8, M: 0.01, TS: 4	F	3052	3049	3054	2897
C: 0.8, M: 0.2, TS: 2	G	3099	2940	2959	2869
C: 0.8, M: 0.2, TS: 4	H	3058	3005	2794	2809

Figure 4.1: List Settings per Population Size

In figure 4.2, the results per population are plotted. The line is corresponds to the mean, while the bars show the spread (min to max) of all 5 repetitions.

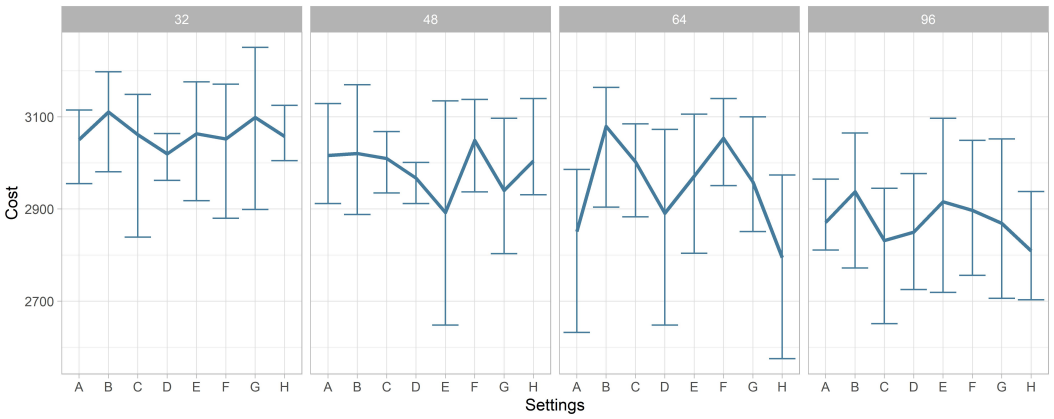

Figure 4.2: mean and error bars per population

A high spread can be seen when looking at small population sizes. Considering these findings, a population size of 96 was chosen. While such a high value will result in a performance impact, it is important to keep the variation low.

4.4 Design of Experiment

Following the conclusion from the previous section 4.3, a population size of 96 will be used. Executing one run for 30 generations currently takes around 3:50 hours. Although two different workstations were available, the time required to execute the needed number of runs for these automated tests would exceed the available time budged. This is without considering a minimum required number of repetitions to remove randomness in the results.

In order to tune the hyperparameter of the genetic algorithm, various different strategies can be used. Using automated hyperparamter tuning approaches like "Grid Search", "Bayesian Optimization, "Simmulated Annealing or "Hyperband" might lead to good results with minimal effort (tuning hyperparameter of these search algorithms is still needed), however they require a high number of runs, which is not feasable. Even a higher level GA can be used find references

" it is not uncommon to automate this process with a simple, top-level “parameter sweep” procedure that systematically adjusts selected parameter values. However, unless care is taken, this can lead to a combinatorial explosion of parameter value combinations and require large amounts of computation time. "K. De Jong, 2007

" The fitness landscapes defined by EA parameter tuning generally have the same unknown properties as the underlying problems that suggested the use of EAs in the first place! So, a natural consequence is the notion of a two-level EA, the top level of which is evolving the parameters of the lower level EA. "K. De Jong, 2007

" My sense is that, for static optimization problems, it will to continue to be the case that particular types of EAs that have been pre-tuned for particular classes of problems will continue to outperform EAs that try to adapt too many things dynamically. "K. De Jong, 2007

A different approach called "design of experiment" (DOE), also known as statistically designed experiments. DOE tries to find the causeand-effect relationship between the factors and the output of experiments. It uses factorial design where each experiment has factors, of which each consists

of at least two settings, with the actual number of settings being called "levels" (Yang and El-Haik, 2009). Design of experiment needs manual expertise to define which factors are possibly of importance and which settings each factor should have, this is a drawback compared to automatic hyperparameter tuning.

"If the range of variable is too small, then we may miss lots of useful information. If the range is too large, then the extreme values might give infeasible experimental runs." (Yang and El-Haik, 2009)

Afterwards, main effects and interactions can be calculated to find the best settings per factor. It provides a graphical representation of these relationship by using interaction as well as main-effects charts. Using ANOVA (Analysis of Variance) it is possible to identify the significance of each factor and interaction, which enables the ranking of these factors. More details on these analysis tools will be provided in section 4.4.3.

A full factorial design will test over all possible combinations of the manually selected factor levels. Looking at the proposed factors in table 4.4.1, we would require 1024 runs[1], which not not feasible performance wise. A full factorial design has the drawback, that as the number if factors k gets increased, the number of needed experimental runs increases exponentially, thus resulting in lengthy experiments. Yang and El-Haik, 2009 state, that most of the results obtained by testing over all combinations are only used for estimating higher-order interactions, which are in most cases insignificant.

"Techniques such as fractional (or partial) factorial experiments are used to simplify the experiment. Fractional factorial experiments investigate only a fraction of all possible combinations. This approach saves considerable time and money but requires rigorous mathematical treatment, both in the design of the experiment and in the analysis of the results. (Roy, 1990)"

[1]number of runs calculated using: `https://datatab.net/statistics-calculator/design-of-experiments`

4.4.1 Taguchi Design

Various improvements to Design of experiment have been but forward by Dr. Genichi Taguchi, such as reducing the influence of uncontrollable (noise) factors on processes and products and reducing variability. Some of these methods evolve around Signal-to-noise (S/N) analysis and utilizing cost functions to "express predicted improvements from DOE results in terms of expected cost saving" (Roy, 1990). This master thesis will not discuss all of Taguchi's proposed considerations, for more detail Roy, 1990 as well as Yang and El-Haik, 2009 is highly recommended.

Find examples of Papers that use taguchi for ga

Using a Taguchi design for evaluating the best hyperparamter has been successfully performed by Dao, Abhary, and Marian, 2016 as well as Assistant Professor, Amity University, Jaipur, Rajasthan, India et al., 2019.

" There are many similarities between "regular" experimental design and Taguchi's experimental design. However, in a Taguchi experiment, only the main effects and two-factor interactions are considered. Higher-order interactions are assumed to be nonexistent. In addition, experimenters are asked to identify which interactions might be significant before conducting the experiment, through their knowledge of the subject matter." (Yang and El-Haik, 2009)

This masters thesis will mainly utilizes Taguchis orthogonal arrays (OAs), "which represent the smallest fractional factorials and are used for most common experiment designs." (Roy, 1990). This means, that only a fraction of combinations needs to be tested which drastically improves performance. Each row of these matrices contains the factors of one experiment, while the columns correspond the factors Hamzaçebi, 2021.

Different orthogonal arrays have been proposed by Taguchi. The researcher has the responsibility to select an array based on the individual needs (Hamzaçebi, 2021). Using these orthogonal arrays instead of full factorial experiments will lead to needing a much smaller amount of simulation runs (in our case only 16 compared to 1024), while the latter "might not provide appreciably more useful information" Roy, 1990.

An orthogonal array has multiple properties: " OFF and OLH experiment designs sample a full factorial design space in a balanced and orthogonal

Definition orthogonal array

fashion. Balance ensures good effect estimates by reducing an estimator's bias and variability. Orthogonality ensures good estimates of two-term interactions. Balance is achieved by ensuring that each level of every factor occurs an equal number of times in the selected sample. Orthogonality is achieved by ensuring that each pair of levels occurs an equal number of times across all experiment parameters in the selected sample. OFF and OLH designs exhibit good space-spanning properties, which aid screening, sensitivity, and comparative analyses. On the other hand, highly fractionated OFF or OLH designs can have poor space-filling properties, which are necessary for optimization analyses.

"Mills, Filliben, and Haines, 2015

As has been stated, probably the biggest drawback of using Taguchi orthogonal arrays is on the one hand to increased manual labour and on the other hand the fact, that higher order interactions ignored.

Roy, 1990 explains why this might not be a big problem: "Generally speaking, OA experiments work well when there is minimal interaction among factors; that is, the factor influences on the measured quality objectives are independent of each other and are linear. In other words, when the outcome is directly proportional to the linear combination of individual factor main effects, OA design identifies the optimum condition and estimates performance at this condition accurately. If, however, the factors interact with each other and influence the outcome, there is still a good chance that the optimum condition will be identified accurately, but the estimate of performance at the optimum can be significantly off. The degree of inaccuracy in performance estimates will depend on the degree of complexity of interactions among all the factors."

This is complimented by Yang and El-Haik, 2009, who states, that: "During many years of applications of factorial design, people have found that higher-order interaction effects (i.e., interaction effects involving three or more factors) are very seldom significant. In most experimental case studies, only some main effects and two-factor interactions are significant."

However, K. De Jong, 2007 claim, that "tuning EA parameters can itself be a challenging task since EA parameters interact in highly non-linear ways."

It remains to be seen in it is possible to use this method for optimizing parameters of a genetic algorithm.

Selection of orthogonal array When choosing a suitable Taguchi orthogonal array, we need to take various factors into account, which can make the process tricky. According to Yang and El-Haik, 2009, we will have to follow a three step procedure:

1. Calculate the total degree of freedom (DOF).
2. Following two rules, standard orthogonal array should be selected:
    a) Total DOF need to be smaller than the number of runs provided by the orthogonal array.
    b) All required factor level combinations need to be accommodated by the orthogonal array.
3. Factors have to be assigned using these rules:
    a) In case the factor level does not fit into the orthogonal array, methods such as column merging and dummy level can be used to modify the original array.
    b) Using the linear graph and interaction table, interactions can be defined.
    c) In case some columns are not assigned, its possible to keep these columns empty.

For this genetic algorithm, 7 factors (3 Factors of Level 4 and 4 Factors of Level 2) have been selected. Which factors to choose and with which level was done based on experience gained on section 4.3. When selecting levels, it is important to have them "as far away from either side of the current working condition as possible."(Roy, 1990) In table 4.4.1, every factor with corresponding levels has been listed,

Using this table, we will now find the best standard orthogonal array in section 4.4.2. Before doing so, it is important to state, that Taguchi allows to test for possible (pre determined) two-level interactions (Yang and El-Haik, 2009). Analysing interactions comes at a cost of Degrees of freedom. If we look at the table, an interaction between ChromosomeType and GeneType might be of interest. Using the power of hindsight, we know, that a second two factor

Factors	Code	Level 1	Level 2	Level 3	Level 4
CrossoverType	A	one point	two point	uniform 0.1	uniform 0.5
CrossoverProp	B	0.2	0.5	0.8	0.9
MutationProp	C	0.01	0.1	0.3	0.5
ChromosomeType	D	Time	Time+NPC	-	-
GeneType	E	int	dict	-	-
TournamentSize	F	2	4	-	-
IndMutationProp	G	0.1	0.5	-	-

Figure 4.3: List of Hyperparamters (Factors) matched to a Code and defined settings (Levels)

interaction is possible within our chosen array, thus we will have a look at the interaction between Tournament Size and IndMutationPropability as well.

4.4.2 Selection of a suitable standart orthogonal array

The total degree of freedom can be quickly calculated using the rules provided by Yang and El-Haik, 2009:

1. 1 DOF is always used for the overall mean.
2. Each factor has a DOF of NumberOfLevels - 1.
3. Two-factor interactions use this equation to calculate DOF: $(n_{factor1} - 1)(n_{factor2} - 1)$ where n = number of levels.

This leads to the following calculation for the needed 3 Factors of Level 4 and 4 Factors of Level 2 as well as the two interactions between ChromosomeType-GeneType and TournamentSize-IndMutationProp:

$$\begin{aligned} DOF &= 1 + 3 * (3 - 1) + 4 * (2 - 1) + 2 * (2 - 1) * (2 - 1) \\ &= 13 \end{aligned} \tag{4.1}$$

A L_{16} array seems suitable to accommodate the required 13 DOF, which can be seen in 4.4.2.

	$L_{16}(2^{15})$														
NO.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
3	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
4	1	1	1	2	2	2	2	2	2	2	2	1	1	1	1
5	1	2	1	1	1	2	2	1	1	2	2	1	1	2	2
6	1	2	2	1	1	2	2	2	2	1	1	2	2	1	1
7	1	2	2	2	2	1	1	1	1	2	2	2	2	1	1
8	1	2	2	2	2	1	1	2	2	1	1	1	1	2	2
9	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
10	2	1	2	1	2	1	2	2	1	2	1	2	1	2	1
11	2	1	2	2	1	2	1	1	2	1	2	2	1	2	1
12	2	1	2	2	1	2	1	2	1	2	1	1	2	1	2
13	2	2	1	1	2	2	1	1	2	2	1	1	2	2	1
14	2	2	1	1	2	2	1	2	1	1	2	2	1	1	2
15	2	2	1	2	1	1	2	1	2	2	1	2	1	1	2
16	2	2	1	2	1	1	2	2	1	1	2	1	2	2	1

Figure 4.4: $L_{16}(2^{15})$ Taguchi ortohogonal array taken from Roy, 1990

This graph now needs to be fitted and modified to accommodate the needed factors. 4 Level Factors need additional space which will be generated using column merging, while interactions will need to be assigned as well. For this, either an interaction table or linear graphs of the L_{16} array can be used (NazanDanacıoğlu, 2005). The linear graph approach is straight forward and will be selected. While there are multiple linear graphs for L_{16} array, 4.4.2 describes the graph which best fits the requirements from table 4.4.1. If no graph with the perfect fit is found, theses graphs can be modified as well, using rules described by NazanDanacıoğlu, 2005.

"In each of Taguchi's orthogonal arrays, there are one or more accompanying linear graphs. A linear graph is used to illustrate the interaction relationships in the orthogonal array."Yang and El-Haik, 2009

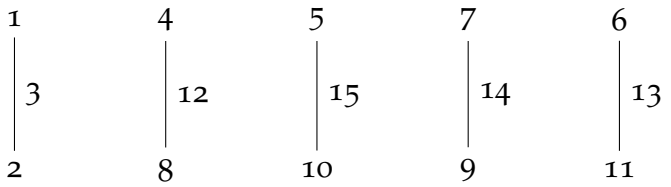

Figure 4.5: Linear Graph of $L_{16}(2^{15})$ taken from Yang and El-Haik, 2009

In a Taguchi linear graph, the nodes as well as the connections both represent columns in the orthogonal array. An interaction between to columns that are represented as nodes "comes out to" to the connecting line column Taguchi et al., 2005. This is useful for both analysing interactions between columns as well as combining (merging) interacting columns in case a higher factor is needed.

Column Merging A, B and C are both 4 level factors. The currently selected orthogonal only fits 2 level factors. Using column merging, it is possible to extend columns to accomodate higher order levels.

As calculated in 4.1, a four-level column requires three degrees of freedom, thus three two-level columns need to be merged. For column merging, it is required, that the to be merged columns are part of an interaction group (Yang and El-Haik, 2009).

So, 3 interaction 2-level columns need to first be selected. One column is discarded, the remain two columns need to be merged using the rules in tabular 4.6.

OLD COLUMN			NEW COLUMN
1	1	->	1
1	2	->	2
2	1	->	3
2	2	->	4

Figure 4.6: Rules taken from Roy, 1990

The four-level factor can then be assigned to this newly generated column. Because three four-level factors are needed for the current experiment, nine two-level columns need to be merged in total.

Assigning Interactions Interactions between two-level factors can be assigned using the linear graph as well. Here, select two connected nodes. The column describing their connection will subsequently contain the interaction (Taguchi et al., 2005).

An interaction between ChromosomeType and GeneType seems possible, thus D and E will be assigned to connected nodes in the linear graph. As we still have some unused space in the graph, we will also look at the interaction of TournamentSize and IndMutationProp (F and G). The resulting graph can be seen in 4.4.2.

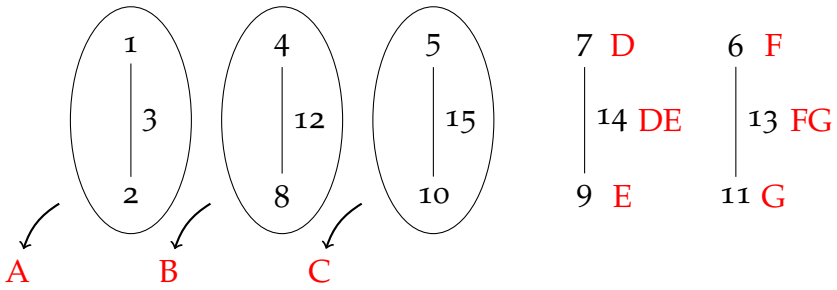

Figure 4.7: Modified Linear Graph to fit our needs

Combining columns 1 2 3 to A, 4 8 12 to B and 5 10 15 to C using rules defined by table 4.6 is done in 4.8.

Removing the old and inserting the new columns in the table and transcoding 7 to D, 9 to E, 14 to DE, 6 to F, 11 to G and 13 to FG results in the final table 4.9. This combinations table will subsequently be used as settings for the simulation runs.

NO.	1 2 ~~3~~	4 8 ~~12~~	5 10 ~~15~~
1	~~1 1~~ > 1	~~1 1~~ > 1	~~1 1~~ > 1
2	~~1 1~~ > 1	~~1 2~~ > 2	~~1 2~~ > 2
3	~~1 1~~ > 1	~~2 1~~ > 3	~~2 1~~ > 3
4	~~1 1~~ > 1	~~2 2~~ > 4	~~2 2~~ > 4
5	~~1 2~~ > 2	~~1 1~~ > 1	~~1 2~~ > 2
6	~~1 2~~ > 2	~~1 2~~ > 2	~~1 1~~ > 1
7	~~1 2~~ > 2	~~2 1~~ > 3	~~2 2~~ > 4
8	~~1 2~~ > 2	~~2 2~~ > 4	~~2 1~~ > 3
9	~~2 1~~ > 3	~~1 1~~ > 1	~~2 1~~ > 3
10	~~2 1~~ > 3	~~1 2~~ > 2	~~2 2~~ > 4
11	~~2 1~~ > 3	~~2 1~~ > 3	~~1 1~~ > 1
12	~~2 2~~ > 3	~~2 2~~ > 4	~~1 2~~ > 2
13	~~2 2~~ > 4	~~1 1~~ > 1	~~2 2~~ > 4
14	~~2 2~~ > 4	~~1 2~~ > 2	~~2 1~~ > 3
15	~~2 2~~ > 4	~~2 1~~ > 3	~~1 2~~ > 2
16	~~2 2~~ > 4	~~2 2~~ > 4	~~1 1~~ > 1

Figure 4.8: Building 4 Level columns from 2 Level columns

4.4.3 Analysing the results

Table 4.9 can now can be used for running all the needed testcases (the interaction columns can be ignored until the evaluation). Transcoding all factors and levels to get the corresponding setting can be done using in the table from 4.4.1. We will repeat every setting 8 times to reduce randomness and gain information about variance. Running the Genetic Algorithm using these 16 different settings each repeated 8 times took 10 days on the two previously described workstations.

ref to section

The results are found in the appendix at 6.2.

Main-effects and interaction chart

Identifying the optimal conditions is done by analyzing the main effects per factor. Using them, it is possible to predict the factors, that lead to the best

NO.	A	B	C	D	E	F	G	FG	DE
1	1	1	1	1	1	1	1	1	1
2	1	2	2	1	2	1	2	2	2
3	1	3	3	2	1	2	1	2	2
4	1	4	4	2	2	2	2	1	1
5	2	1	2	2	1	2	2	1	2
6	2	2	1	2	2	2	1	2	1
7	2	3	4	1	1	1	2	2	1
8	2	4	3	1	2	1	1	1	2
9	3	1	3	2	2	1	2	2	1
10	3	2	4	2	1	1	1	1	2
11	3	3	1	1	2	2	2	1	2
12	3	4	2	1	1	2	1	2	1
13	4	1	4	1	2	2	1	2	2
14	4	2	3	1	1	2	2	1	1
15	4	3	2	2	2	1	1	1	1
16	4	4	1	2	1	1	2	2	2

Figure 4.9: Final version of used Taguchi orthogonal array

result Roy, 1990.

Yang and El-Haik, 2009 explains them well: “The main-effects chart is a plot of average responses at different levels of a factor versus the factor levels”

So, for every factor, sum up the mean of all results per level, then divide by the number of runs per level.

Example for D

The resulting main-effect charts can be seen here:

In case there is no interaction, the optimal setting is easily determined by using the main effects chart. Go over every factor in the chart and use the best level (in case of this experiment, the level with the lowest cost value). If interactions exist, they might have an influence on the best settings and need to be investigated (Yang and El-Haik, 2009).

To investigate previously defined interactions, a test of interactions can be used. Their calculation is similar to calculating main effects.

Example for DE

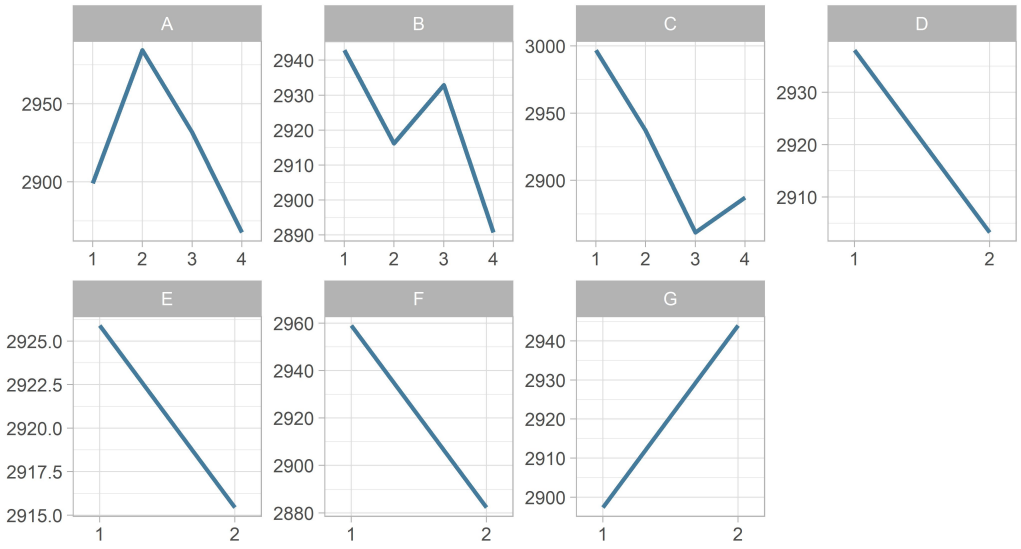

Figure 4.10: Main Effects

If lines cross, an interaction between the two factors exists. The more parallel the lines are, the less likely an interaction. Magnitude of the angle between the lines corresponds to the degree of interaction presence, according to Roy, 1990.

ANOVA

Before choosing the best settings, ANOVA analysis (analysis of variance) should be performend on the results. Among other things, this will provide information on the magnitude of contribution of each main effects and interactions. The calculation of ANOVA is the same as for a classical design of experiment, according to Yang and El-Haik, 2009.

Summary after reading "Introduction into R"

"In analysis of variance, mean squares are used in the F test to see if the corresponding effect is statistically significant."Yang and El-Haik, 2009 "F ratio is a better measure for relative performance"Yang and El-Haik, 2009

". The most commonly used criterion is to compare the p value with 0.05,

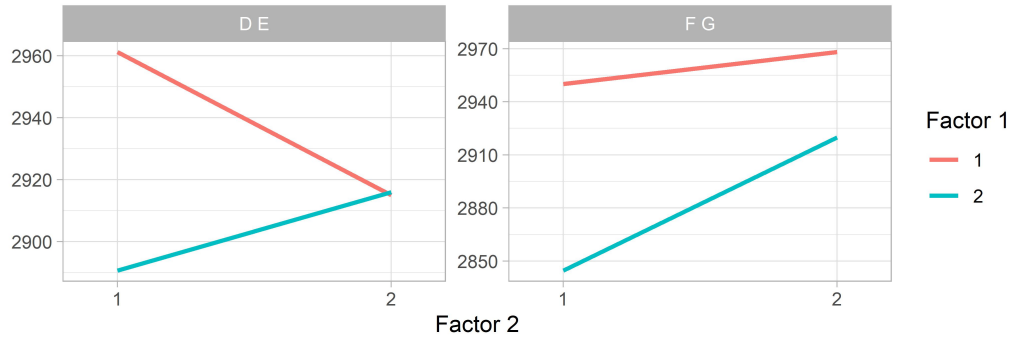

Figure 4.11: Test of interactions

or 5%, if p value is less than 0.05, then that effect is significant."Yang and El-Haik, 2009

"The variance ratio, commonly called the F statistic, is the ratio of variance due to the effect of a factor and variance due to the error term. (The F statistic is named after Sir Ronald A. Fisher.) This ratio is used to measure the significance of the factor under investigation with respect to the variance of all of the factors included in the error term. The F value obtained in the analysis is compared with a value from standard F-tables for a given statistical level of significance." Roy, 1990.

Due to our number of repetitions, the number of DOF increases according to the following equation (taken from Roy, 1990):

$$\begin{aligned} DOF &= totalNumberOfResults - 1 \\ &= numberOfTrials * numberOfRepetitions - 1 \\ &= 16 * 8 - 1 = 127 \end{aligned} \tag{4.2}$$

Calculating ANOVA can be done simply be done using R, which will result in table 4.1.

A, C, F and G have a relatively high F value, which suggests high influence on the model.

explain using R book

The Multiple R-squared: 0.4275, Adjusted R-squared: 0.3509 ... both are bad.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	3	238901.41	79633.80	6.66	0.0004
B	3	49972.09	16657.36	1.39	0.2488
C	3	343169.03	114389.68	9.56	0.0000
D	1	38781.12	38781.12	3.24	0.0745
E	1	3507.03	3507.03	0.29	0.5893
F	1	189112.50	189112.50	15.81	0.0001
G	1	69751.13	69751.13	5.83	0.0174
D:E	1	41041.12	41041.12	3.43	0.0666
F:G	1	26277.78	26277.78	2.20	0.1411
Residuals	112	1339693.00	11961.54		

Table 4.1: ANOVA results

We can also look at the percentage contribution of each factor, using the formula gathered by Yang and El-Haik, 2009:

$$SS_T = SS_A + SS_B + SS_C + ... + SS_{error} \tag{4.3}$$

$$contribution_A = SS_A / SS_T * 100 \tag{4.4}$$

The percentage contribution is plotted in 4.4.3 (Sum of all factor contributions == Multiple R-squared in theory)

We can clearly see a high contribution of the residuals (error), which is concerning.

Selection of optimal setting

When choosing the optimal setting, the first step is to look at the best main effects combination. For this experiment, the best combination would be the following: A4, B4, C3, D2, E2, F2, G1. Looking the ANOVA table,

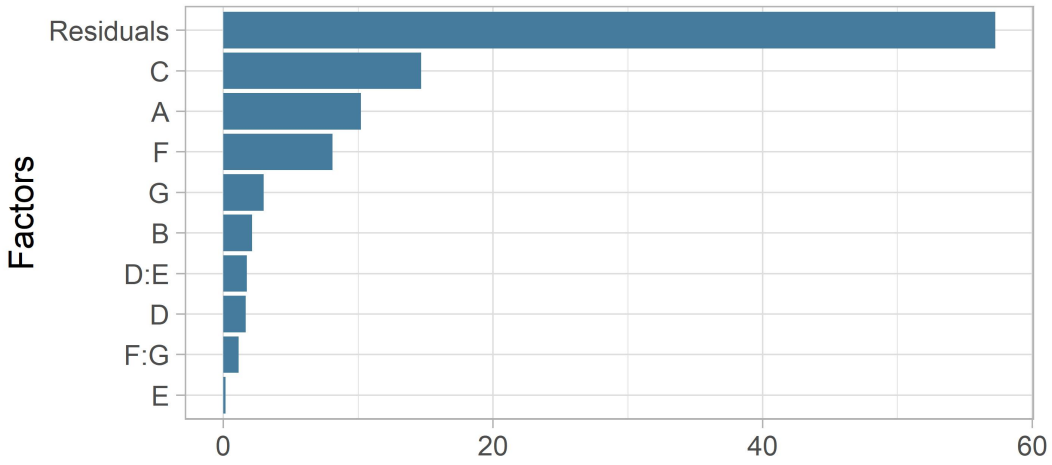

Figure 4.12: Percentage Contribution

the interaction D:E has seams to have significance, especially compared to E. We will try to integrate this interaction. Compared to D:E, the second interaction (F:G) has a lower influence. This is also the case when looking at the individual factors F and G. The interaction F:G will thus not further be discussed.

The test of interaction in figure 4.4.3 suggest D2 and E1 as the best combination. This is optimal, as D2 is also suggested by the main effects. E1 is different to the suggested main effects, however its low F value in the anova table suggests low significance for using E2. Concluding this line of thought, the combination A4, B4, C3, D2, E1, F2, G1 looks to be optimal.

Optimum performance calculation Using optimal performance calculation 1. using only main effects or 2. using main effects with applied interaction can be applied. The equation provided by Roy, 1990.

Better bars

$$\begin{aligned} Y_{opt} &= \bar{T} + (\bar{A}_4 - \bar{T}) + (\bar{B}_4 - \bar{T}) + (\bar{C}_3 - \bar{T}) + (\bar{D}_2 - \bar{T}) + (\bar{E}_2 - \bar{T}) + (\bar{F}_2 - \bar{T}) + (\bar{G}_1 - \bar{T}) \\ &= 2693.984 \end{aligned} \tag{4.5}$$

$$\begin{aligned} Y_{opt} &= \bar{T} + (\bar{A}_4 - \bar{T}) + (\bar{B}_4 - \bar{T}) + (\bar{C}_3 - \bar{T}) + (\bar{D}_2 - \bar{T}) + (\bar{E}_1 - \bar{T}) + ([\bar{Dx}E]_2 - \bar{T}) + (\bar{F}_2 \\ &= 2686.547 \end{aligned} \tag{4.6}$$

Using the interaction D:E, the performance estimation improves from 2693.984 to 2686.547. Considering this result, interaction will be taken into account and the optimized settings are as follows: CrossoverType: Uniform 0.5, CrossoverPropability: 0.9, MutationPropability: 0.3, ChromosomeType: Time+NPC, GeneType: integer encoding, TournamentSize: 4 and Individual-MutationPropability: 0.1.

signal-to-noise (S/N) As previously discussed when looking at the anova model, the error is very high, which suggests high randomness. Taguchi recommends using signal-to-noise (S/N) ratio to reduce the variability, as using only the mean of the results does not take the variation into account (Roy, 1990). The greater the signal-to-noise ratio, the smaller the variance. Roy, 1990 further states, that the "use of the S/N ratio offers an objective way to look at the two characteristics (consistency and average value) together."

When using S/N, todo: talk about equation.

Afterwards, generating main effects and anova table is the same as using the mean. However the DOF calcualtion done in equation 4.2 is no longer valid, as repetitions get combined to 1 value per run. So, the DOF for anova changes to the following equation in 4.7 according to Roy, 1990.

$$\begin{aligned} DOF &= totalNumberOfResults - 1 \\ &= numberOfTrials * 1 - 1 \\ &= 16 - 1 = 15 \end{aligned} \tag{4.7}$$

Find out why 15 DOF is not enough

This is not enough residuals for generating the F value in anova. Thus in order to reduce the current DOF, the anova table was generated without the interaction F:G considered, which can be seen in 4.2.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	3	0.26	0.09	3.01	0.3953
B	3	0.05	0.02	0.60	0.7139
C	3	0.38	0.13	4.44	0.3326
D	1	0.04	0.04	1.48	0.4378
E	1	0.00	0.00	0.12	0.7845
F	1	0.21	0.21	7.35	0.2250
G	1	0.08	0.08	2.80	0.3429
D:E	1	0.04	0.04	1.56	0.4296
Residuals	1	0.03	0.03		

Table 4.2: S/N ANOVA results

When looking at the p values of this table, it is very obvious that no factor can discard the null - hypothesis, which states that a factor has no significant effect. Considering this, it was deemed to be not necessary to perform further investigations. Somehow argument, that having less variablity is only to an extend important. It is always possible to restart a simulation, if the variability is not too large, it is important that the results have a good mean overall. Less variability is in producing products much more important.

Elite Although the optimal hyperparameter setting will be discussed in chapter 5, a problem was obvious when analyzing a run using the optimzed GA. Figure 4.4.3 shows for a few selected repetitions the best individual cost per generation.

The lines show that setbacks in the optimal cost between two generations happens frequenctly. In order to mitigate this problem, it was decided to implement elite selection with a size of 2. This means, that per generation, the two best individuals are copied into the next generation without modifications, which makes worse performance between generations not possible. It is important to note, that the two best individuals can still be selected by tournament selection for modification, its just that a copy of them is saved. Figure **??** shows the effect of these changes.

Comparing the 10 repetitions also provides a clear picture in figure 4.4.3. It is thus concluded that the slightly modified version of the optimized Ga

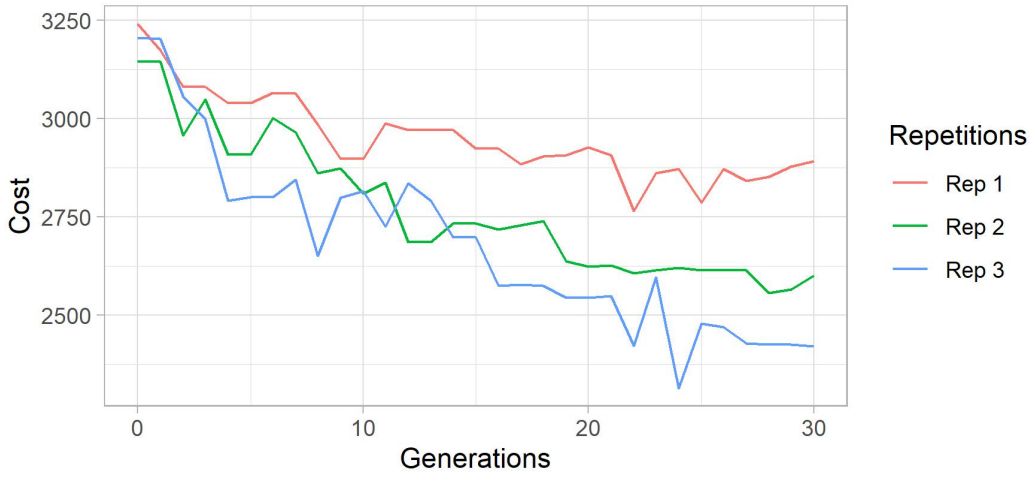

Figure 4.13: Genetic Algorithm without Elite

now using Elite of 2 will be used for chapter 5.

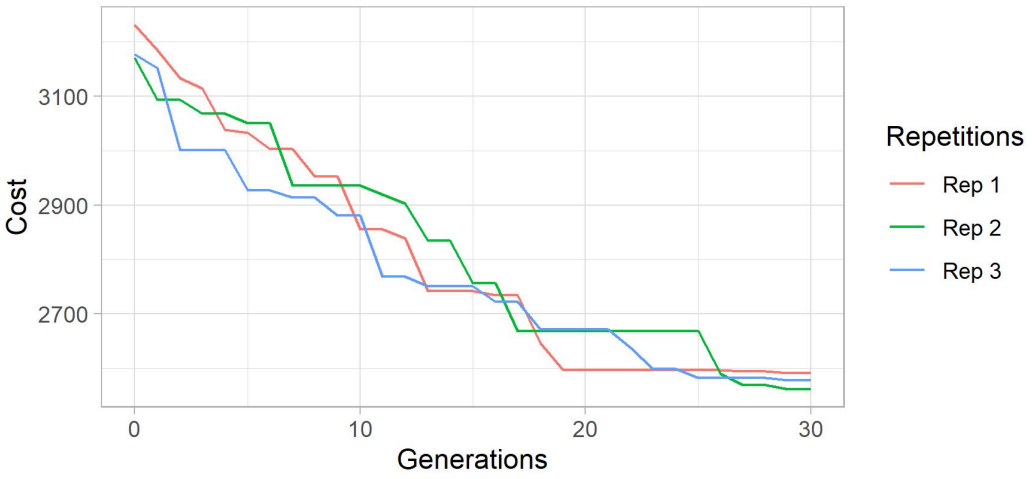

Figure 4.14: Genetic Algorithm with Elite

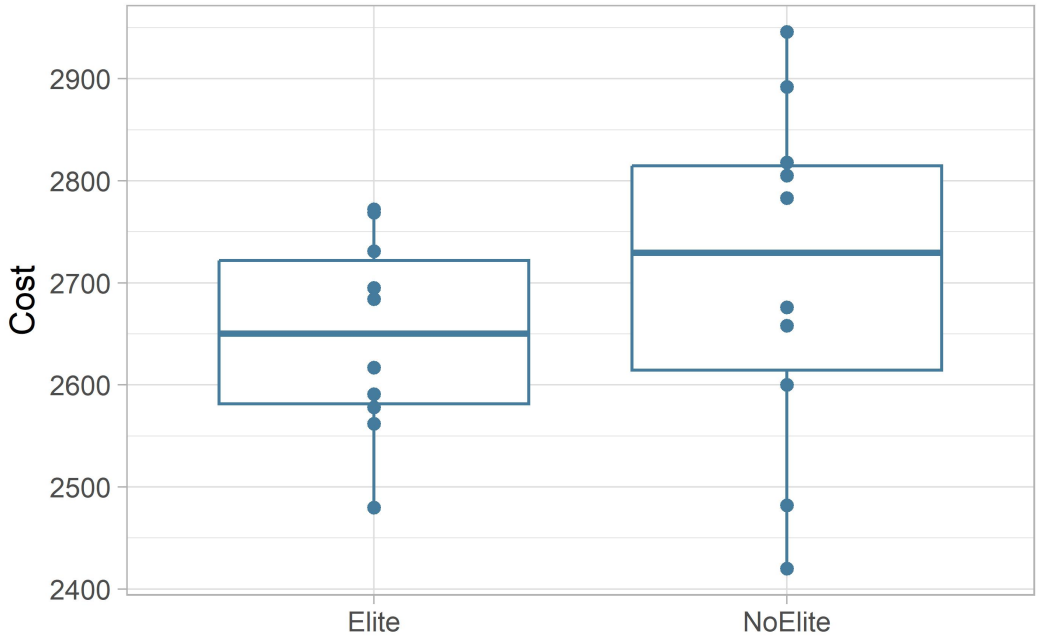

Figure 4.15: Comparison Elite vs No Elite

5 Evaluation

in this chapter, we evaluate and compare various different settings

"We found crossover and mutation most influential in GA success."Mills, Filliben, and Haines, 2015

"The cross over operator is found to be the most influential parameter in both the case studies, followed by mutation rate, population size for case study-1 and population size and selection process for case study-2. It is evident that the robust GA parameter settings are sensitive"Majumdar and Ghosh, 2015

Boyabatli and Sabuncuoglu, 2004 also suggest a high mutation rate.

It seams like the high variations does not allow for more complex traffic situations, giving a clear priority to pedestrian actions

TODO: chart: emergency break due to pedestrians vs vehicles

5.1 Comparison with random and default ga Values

scenario 1: default : 9v 5p

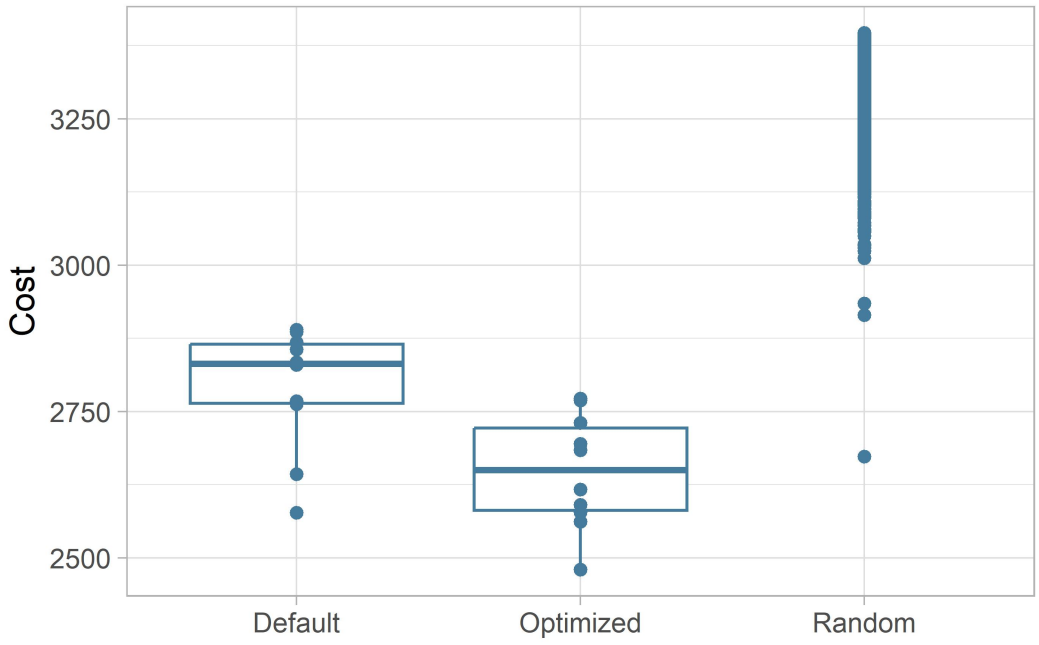

Figure 5.1: Genetic Algorithm with Elite

5.2 Generalization on different start scenarios

Scenario 2

scenario 2: 9v 5p

Scenario 3

scenario 3: 5v 3p

Scenario 4

scenario 4: 18v 10p

also compare (average) diversity?

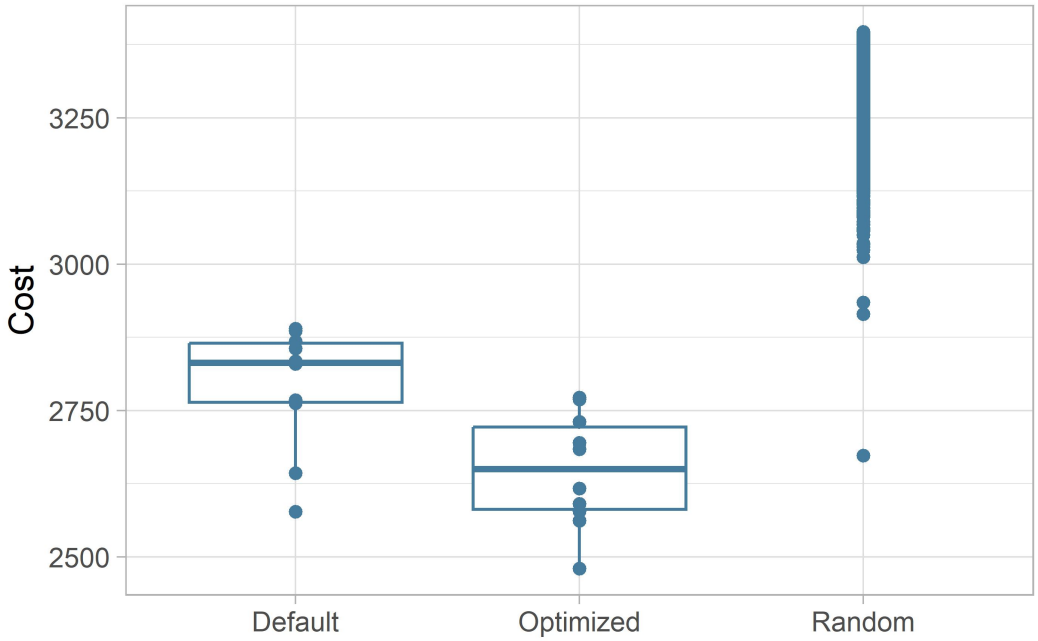

Figure 5.2: Genetic Algorithm with Elite

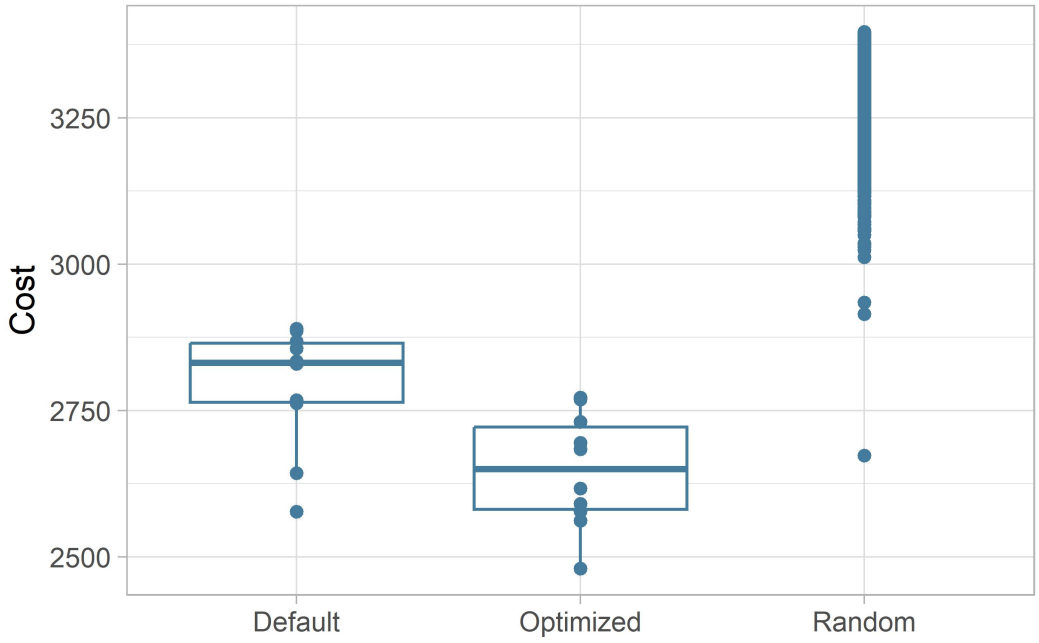

Figure 5.3: Genetic Algorithm with Elite

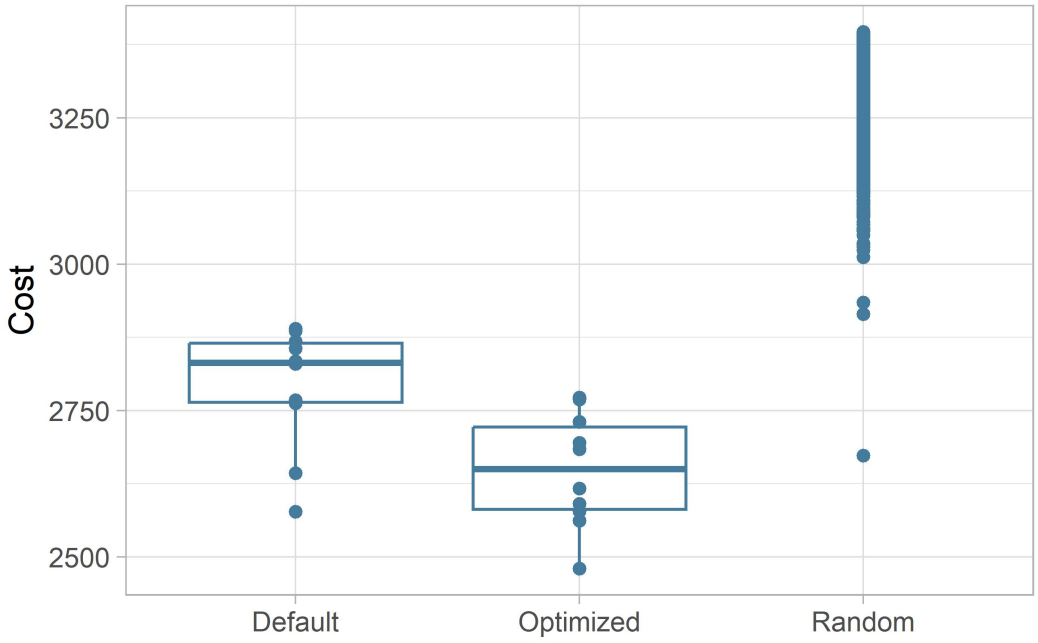

Figure 5.4: Genetic Algorithm with Elite

6 Conclusion

6.1 Future Work

6.1.1 Oracles

While not implemented here, Oracles are needed in order to get a list of good scenarios.

6.2 Final words

Appendix

Appendix A.

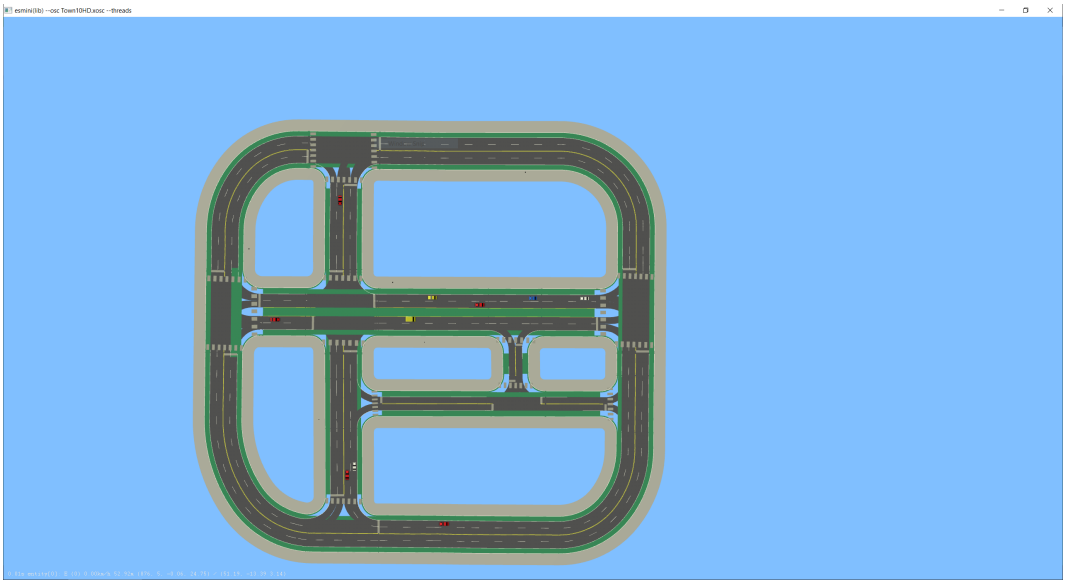

Figure 1: Start scenario 1

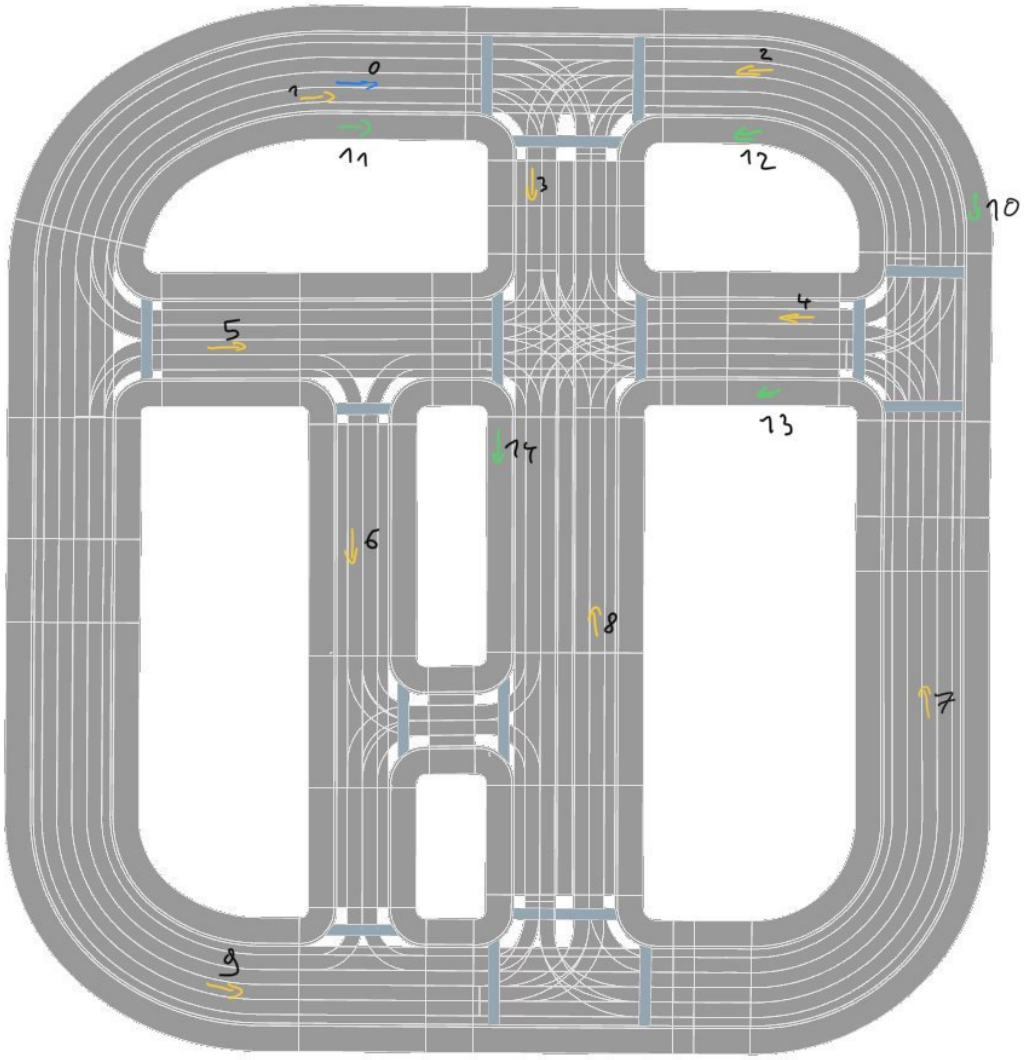

Figure 2: Start scenario 2

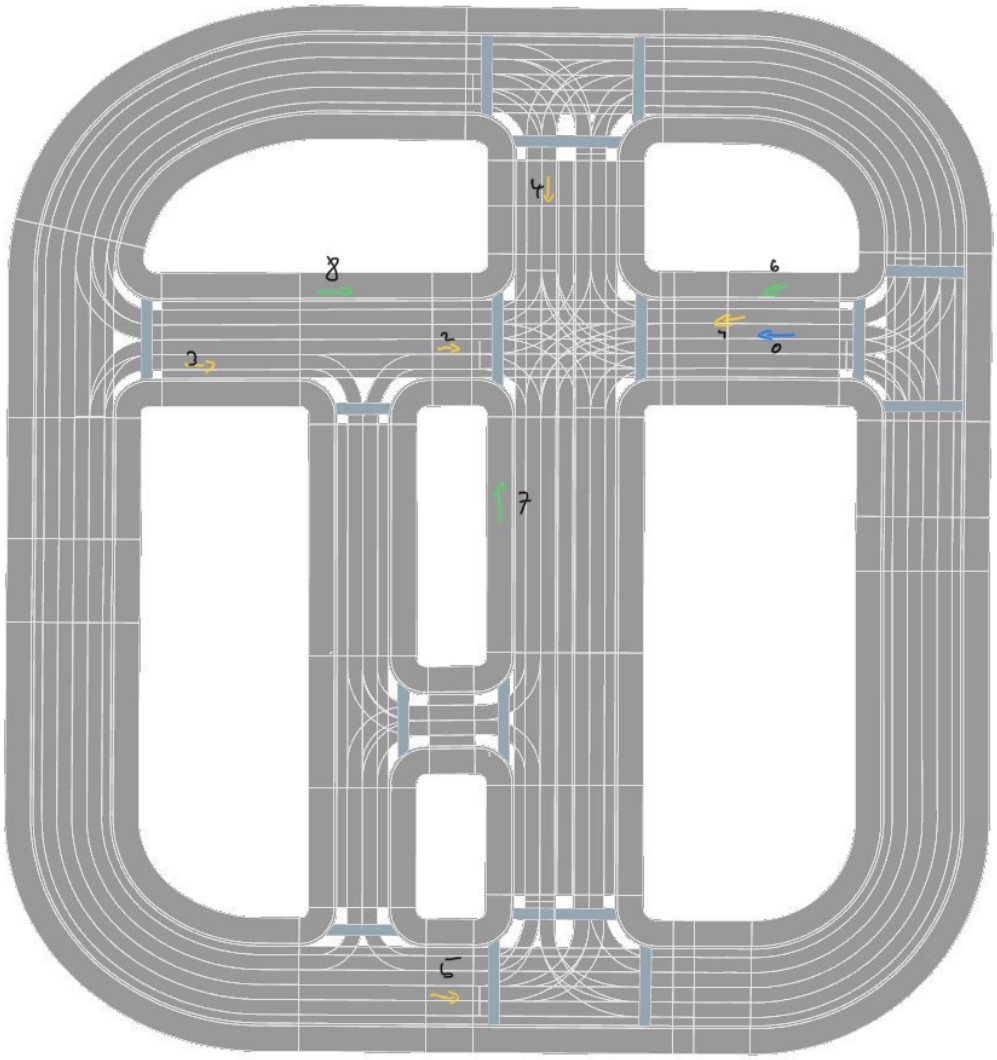

Figure 3: Start scenario 3

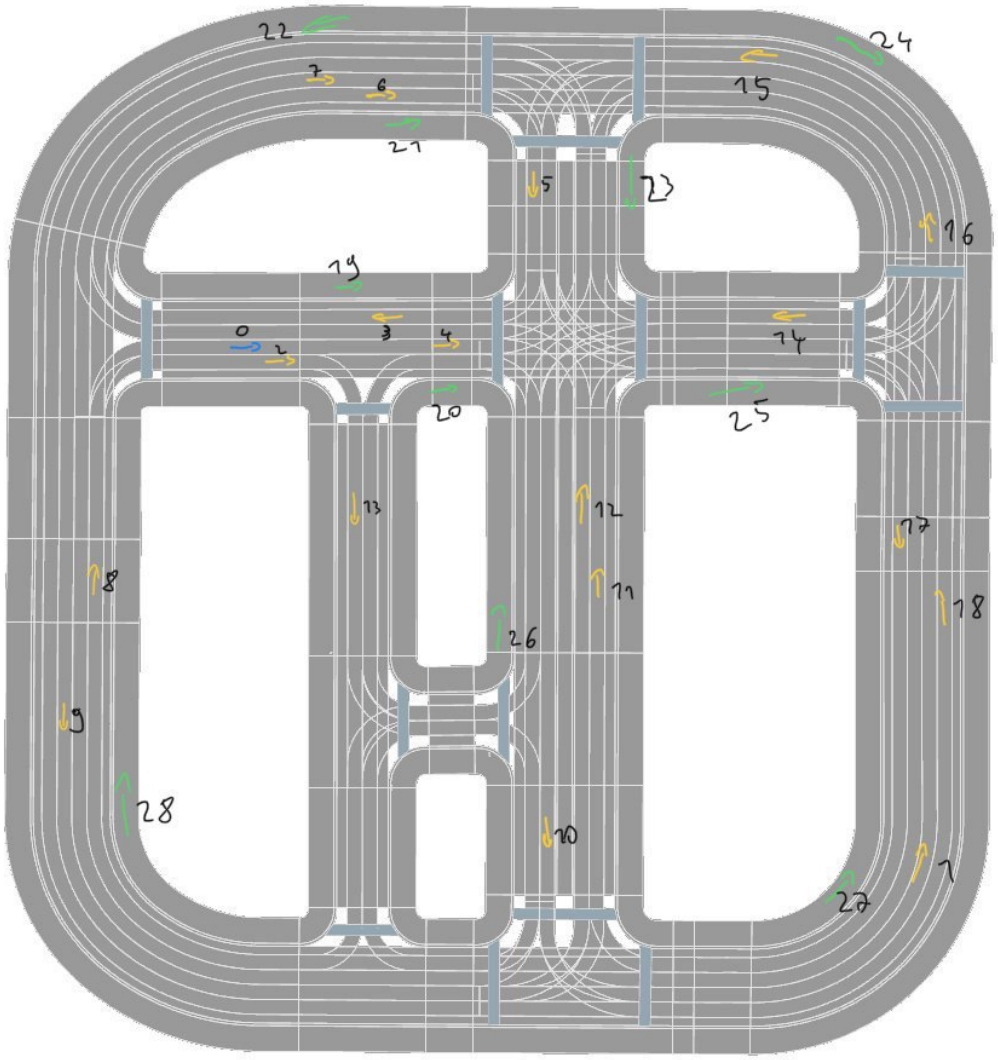

Figure 4: Start scenario 4

Appendix B.

NO.	rep1	rep2	rep3	rep4	rep5	rep6	rep7	rep8
1	3124	3110	3025	3077	3068	2925	3106	3105
2	2694	2980	3025	2996	3037	2921	3068	2900
3	2638	2711	2624	2856	2623	2832	2904	2778
4	2735	2805	2851	2965	2876	2703	2848	2858
5	3074	2955	3045	3080	3120	2971	2895	2798
6	2974	2979	2929	2941	2952	2936	3139	2953
7	3108	3099	3049	3020	3092	3057	3090	2895
8	2840	2931	2921	2921	2957	2997	2889	2895
9	3007	2995	2983	3009	2847	2996	2734	2927
10	2916	2828	3013	2787	2818	2926	3034	2822
11	3007	2879	3090	3033	2906	2981	3109	3104
12	2946	3016	2790	2917	2904	2983	2898	2606
13	2378	2712	2906	2800	2912	2795	2860	2834
14	2895	2760	2750	2849	2542	2997	2965	2991
15	2842	3065	3050	2779	2862	2923	2955	2892
16	3065	2834	2643	3056	3051	3011	2828	2963

Figure 1: List of results

Appendix C.

Insert information of Gene action probabitlities

Bibliography

Almanee, Sumaya et al. (2021). "scenoRITA: Generating Less-Redundant, Safety-Critical and Motion Sickness-Inducing Scenarios for Autonomous Vehicles." In: Publisher: arXiv Version Number: 1. DOI: 10.48550/ARXIV.2112.09725. URL: https://arxiv.org/abs/2112.09725 (visited on 10/19/2023) (cit. on p. 34).

Assistant Professor, Amity University, Jaipur, Rajasthan, India et al. (July 30, 2019). "Parameter Tuning Method for Genetic Algorithm using Taguchi Orthogonal Array for Non-linear Multimodal Optimization Problem." In: *International Journal of Recent Technology and Engineering (IJRTE)* 8.2, pp. 2979–2986. ISSN: 22773878. DOI: 10.35940/ijrte.B2711.078219. URL: https://www.ijrte.org/portfolio-item/B2711078219/ (visited on 10/09/2023) (cit. on pp. 35, 39).

Boyabatli, Onur and Ihsan Sabuncuoglu (2004). "Parameter selection in genetic algorithms." In: *Journal of Systemics, Cybernetics and Informatics* 4.2. Publisher: International Institute of Informatics and Cybernetics, p. 78 (cit. on pp. 4, 34, 57).

Dao, Son, Kazem Abhary, and Romeo Marian (Feb. 2016). "Maximising Performance of Genetic Algorithm Solver in Matlab." In: *Engineering Letters* 24 (cit. on pp. 31, 35, 39).

De Jong, Kenneth (2007). "Parameter Setting in EAs: a 30 Year Perspective." In: *Parameter Setting in Evolutionary Algorithms*. Ed. by Fernando G. Lobo, Cláudio F. Lima, and Zbigniew Michalewicz. Red. by Janusz Kacprzyk. Vol. 54. Series Title: Studies in Computational Intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–18. ISBN: 978-3-540-69431-1 978-3-540-69432-8. DOI: 10.1007/978-3-540-69432-8_1. URL: http://link.springer.com/10.1007/978-3-540-69432-8_1 (visited on 10/13/2023) (cit. on pp. 5, 7, 13, 17, 31–34, 37, 40).

De Jong, Kenneth Alan (1975). *An analysis of the behavior of a class of genetic adaptive systems.* University of Michigan (cit. on p. 34).

Erickson, Mark, Alex Mayer, and Jeffrey Horn (Jan. 2002). "Multi-objective optimal design of groundwater remediation systems: application of the niched Pareto genetic algorithm (NPGA)." In: *Advances in Water Resources* 25.1, pp. 51–65. ISSN: 03091708. DOI: 10.1016/S0309-1708(01)00020-3. URL: https://linkinghub.elsevier.com/retrieve/pii/S0309170801000203 (visited on 10/23/2023) (cit. on p. 19).

Fazal, M.A. et al. (Mar. 2005). "Estimating groundwater recharge using the SMAR conceptual model calibrated by genetic algorithm." In: *Journal of Hydrology* 303.1, pp. 56–78. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2004.08.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S0022169404003890 (visited on 10/23/2023) (cit. on p. 35).

Grefenstette, John (Jan. 1986). "Optimization of Control Parameters for Genetic Algorithms." In: *IEEE Transactions on Systems, Man, and Cybernetics* 16.1, pp. 122–128. ISSN: 0018-9472. DOI: 10.1109/TSMC.1986.289288. URL: http://ieeexplore.ieee.org/document/4075583/ (visited on 10/13/2023) (cit. on pp. 3–5, 7, 12–14, 33, 34).

Hamzaçebi, Coşkun (Mar. 24, 2021). "Taguchi Method as a Robust Design Tool." In: *Quality Control - Intelligent Manufacturing, Robust Design and Charts*. Ed. by Pengzhong Li, Paulo António Rodrigues Pereira, and Helena Navas. IntechOpen. ISBN: 978-1-83962-497-1 978-1-83962-498-8. DOI: 10.5772/intechopen.94908. URL: https://www.intechopen.com/books/quality-control-intelligent-manufacturing-robust-design-and-charts/taguchi-method-as-a-robust-design-tool (visited on 10/28/2023) (cit. on p. 39).

Hussain, Abid and Yousaf Shad Muhammad (Apr. 2020). "Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator." In: *Complex & Intelligent Systems* 6.1, pp. 1–14. ISSN: 2199-4536, 2198-6053. DOI: 10.1007/s40747-019-0102-7. URL: http://link.springer.com/10.1007/s40747-019-0102-7 (visited on 07/23/2023) (cit. on pp. 4, 5, 7–11).

Jinghui Zhong et al. (2005). "Comparison of Performance between Different Selection Strategies on Simple Genetic Algorithms." In: *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*. International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies

and Internet Commerce (CIMCA-IAWTIC'06). Vol. 2. Vienna, Austria: IEEE, pp. 1115–1121. ISBN: 978-0-7695-2504-4. DOI: 10.1109/CIMCA.2005.1631619. URL: http://ieeexplore.ieee.org/document/1631619/ (visited on 10/23/2023) (cit. on pp. 9, 10, 34, 35).

Katoch, Sourabh, Sumit Singh Chauhan, and Vijay Kumar (Feb. 1, 2021). "A review on genetic algorithm: past, present, and future." In: *Multimedia Tools and Applications* 80.5, pp. 8091–8126. ISSN: 1573-7721. DOI: 10.1007/s11042-020-10139-6. URL: https://doi.org/10.1007/s11042-020-10139-6 (visited on 03/28/2023) (cit. on pp. 4–9, 11–13, 15, 18).

Klampfl, Lorenz, Florian Klück, and Franz Wotawa (2023). "Using genetic algorithms for automating automated lane-keeping system testing." In: *Journal of Software: Evolution and Process* n/a (n/a). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/smr.2520, e2520. ISSN: 2047-7481. DOI: 10.1002/smr.2520. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/smr.2520 (visited on 03/29/2023) (cit. on pp. 14, 17, 32).

Majumdar, Abhishek and Debashis Ghosh (Oct. 15, 2015). "Genetic Algorithm Parameter Optimization using Taguchi Robust Design for Multiresponse Optimization of Experimental and Historical Data." In: *International Journal of Computer Applications* 127.5, pp. 26–32. ISSN: 09758887. DOI: 10.5120/ijca2015906383. URL: http://www.ijcaonline.org/research/volume127/number5/majumdar-2015-ijca-906383.pdf (visited on 10/27/2023) (cit. on pp. 4, 7, 14, 18, 19, 57).

Marsili Libelli, S. and P. Alba (July 2000). "Adaptive mutation in genetic algorithms." In: *Soft Computing* 4.2, pp. 76–80. ISSN: 1432-7643. DOI: 10.1007/s005000000042. URL: http://link.springer.com/10.1007/s005000000042 (visited on 11/21/2023) (cit. on pp. 4, 15, 16).

Mills, K. L., J. J. Filliben, and A. L. Haines (June 2015). "Determining Relative Importance and Effective Settings for Genetic Algorithm Control Parameters." In: *Evolutionary Computation* 23.2, pp. 309–342. ISSN: 1063-6560, 1530-9304. DOI: 10.1162/EVCO_a_00137. URL: https://direct.mit.edu/evco/article/23/2/309-342/986 (visited on 10/13/2023) (cit. on pp. 3, 4, 7, 16, 33–35, 40, 57).

NazanDanacıoğlu, F. ZehraMuluk (2005). "TAGUCHI TECHNIQUES FOR 2^k FRACTIONAL FACTORIAL EXPERIMENTS." In: *Journal* 34.1, pp. 83–93. ISSN: 2651-477X-2651-477X (cit. on p. 43).

Roy, Ranjit K. (1990). *A primer on the Taguchi method*. Competitive manufacturing series. New York: Van Nostrand Reinhold. 247 pp. ISBN: 978-0-442-23729-5 (cit. on pp. 1, 38–41, 43, 44, 47–49, 51, 52).

Srinivas, M. and L.M. Patnaik (June 1994). "Genetic algorithms: a survey." In: *Computer* 27.6, pp. 17–26. ISSN: 0018-9162. DOI: 10.1109/2.294849. URL: http://ieeexplore.ieee.org/document/294849/ (visited on 10/13/2023) (cit. on pp. 3, 4, 6–8, 12–14, 17, 34).

Taguchi, Genichi et al. (2005). *Taguchi's quality engineering handbook*. Hoboken, N.J. : Livonia, Mich: John Wiley & Sons ; ASI Consulting Group. 1662 pp. ISBN: 978-0-471-41334-9 (cit. on pp. 44, 45).

Whitley, Darrell (June 1, 1994). "A genetic algorithm tutorial." In: *Statistics and Computing* 4.2, pp. 65–85. ISSN: 1573-1375. DOI: 10.1007/BF00175354. URL: https://doi.org/10.1007/BF00175354 (visited on 03/28/2023) (cit. on p. 34).

Xia, Xuemin et al. (Feb. 1, 2019). "Genetic algorithm hyper-parameter optimization using Taguchi design for groundwater pollution source identification." In: *Water Supply* 19.1, pp. 137–146. ISSN: 1606-9749, 1607-0798. DOI: 10.2166/ws.2018.059. URL: https://iwaponline.com/ws/article/19/1/137/39199/Genetic-algorithm-hyperparameter-optimization (visited on 10/06/2023) (cit. on p. 4).

Yang, Kai and Basem El-Haik (2009). *Design for six sigma: a roadmap for product development*. 2nd ed. London: McGraw-Hill [distributor]. 741 pp. ISBN: 978-0-07-154767-3 (cit. on pp. 38–44, 47–50).