The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Department of Civil Engineering

AN APPROACH TO THE HIGHWAY ALIGNMENT DEVELOPMENT PROCESS USING GENETIC ALGORITHM BASED OPTIMISATION

by

Botan Majeed Ahmad AL-Hadad

Thesis submitted to the University of Nottingham for
the degree of Doctor of Philosophy

October 2011

ABSTRACT

Highway alignment development is recognised as a non-linear constrained optimisation problem. It is affected by many economical, social, and environmental factors subject to many design constraints. The highway alignment development process is therefore considered complex but very important.

Highway alignment development is about finding an optimum alignment solution between two termini points in a 3D space, subject to several constraints. The development process using the current method is considered complex because of the number of the design elements involved, their interactions, and the formulations required to relate them to a realistic highway alignment. Moreover, it is considered that an alignment, generated using the existing method, results in a sub-optimal solution. This is due to the fact that the two alignments (horizontal and vertical alignments) are found in two independent stages and from only a handful number of alternative evaluations.

This research reports on a new approach for improving the process of highway alignment development by utilising modern technologies. It proposes a novel design approach, as an alternative to the existing method, for highway alignment development in a three-dimensional space (considering the horizontal and vertical alignments simultaneously). It describes a method for highway alignment development through *station points*. Station points, as points along the centre line of alignment which are defined by their X, Y, and Z coordinates, are used to define the alignment configuration. The research also considers the implications of access provision (in term of junctions) and their locations on highway alignment. The environmental factors (noise and air pollution in terms of proximity distance) and accessibility (user and link construction costs in terms of access costs) are embedded in the formulations required to represent junctions in the model.

The proposed approach was tested through the development of a genetic algorithms based optimisation model. To achieve this, several algorithms were developed to perform the search. The evaluation of the solutions was handled by a fitness function that includes construction (length), location (land acquisition, environmentally sensitive areas, and soil condition), and earthwork (fill and cut material) dependent costs. Other forms of costs that are quantifiable can also be incorporated within the fitness function. The critical constraints, believed important for realistic alignments (horizontal curvature, vertical curvature, and maximum gradient) are also dealt with within the model formulation.

The experimental results show that the problem of highway alignment can be better represented using the concept of station points, by which better alignment solutions (global or near global solutions) were achieved. It was also shown that the alignment development process could be simplified through the use of station points, resulting in the efficient evaluation of more alternatives. Furthermore, the results conclude that a highway alignment cannot be optimum unless it is simultaneously optimised with junctions. Further investigations and development are also recommended for future studies.

PUBLISHED PAPERS

The following papers have been published during this research study period:

AL-Hadad, B., Mawdesley, M., & Stace, R., (2010). A Genetic Algorithm Approach to Highway Alignment Development, *Proceedings of the International Conference on Computing in Civil and Building Engineering (icccbe)*, Nottingham, UK.

AL-Hadad, B. & Mawdesley, M., (2010). A Genetic Algorithm Approach to a 3D Highway Alignment Development, *Proceedings of the International Conference on Evolutionary Computation (ICEC)*, Valencia, Spain.

ACKNOWLEDGMENTS

I would like to give my grateful appreciation to my supervisor, Dr Michael Mowdesley, for his continuous and generous support. I wonder how a man with all his business commitments is able to schedule weekly meetings with his students. His supervision and encouragement were the key reasons for the success of this work. He did not leave us without his support, even when he was away and on holidays. He kept in contact via the internet and Skype meetings. He was not only a supervisor, but also an inspiration behind the ideas and the arguments to complete the work. His humble and heart-warming personality made me confident and proud of myself. Great thanks to him.

I also extend my thanks to the University of Nottingham for providing the required facilities and making an excellent environment to study in. The feeling of being embraced by nature is what Nottingham University Park provides.

I wish also to deeply thank my parents, sister, and brother. Whenever I needed them, they were handy to relieve my pains and purify my soul. Their continuous help cannot be forgotten.

Finally, special thanks to my beloved wife, Fenk. She made me wonder at how people can be that patient. She sacrificed her emotions and time just to make me feel comfortable and happy with my study. Her endless love and support cannot be quantified. A great love to my daughter, Sawa, and son, Alend as well. While I was at home studying, their freedom was restricted so as not to disturb me. Like adults, they were able to understand what I had to do to achieve my goals. I owe them a great debt.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

This thesis describes a research which was done to investigate how to determine the best alignment for a highway. It is important because the development of socio-economic generation of any area is linked with the availability of a road system that is optimally positioned and designed alongside the facilities that the alignment should provide (efficient access) to the land uses of the area under consideration.

Highway alignment development, through a repetitve process, aims to connect two terminal points at minimum possible cost (economical path) subject to the design, environmental, economical, social, and political constraints (Jha et al, 2007; Kim et al, 2003). Highway alignment is part of a wider system of a road network; therefore, considering access to the different land uses, while preserving the environment of the area, may greatly promote socio-economic generation and guarantees a sustainable development. Air pollution and damage to the land features are two main environmental concerns with the highway alignment problem. Studies showed that 60% of CO emission in the big cities comes from road traffic (Yang et al, 2003). Developing a highway alignment must also consider the safety of the users. This can be achieved by adhering to the geometric design policies of the relevant national and/or international design standards of the road type.

The problem, according to the purpose that the alignment is built for, may also take values, as input, from the planner/designer, public, interested groups, politicians, and decision makers. These inputs, in some ways, may shape the design, location, and cost of the solution. These and many other components (e.g. maintenance costs) can also be included in the highway alignment development process.

All these parameters must be considered together. However, an alignment that makes one of the parameters optimal will rarely, if ever, make them all optimal. The problem of developing and selecting an optimum alignment is therefore very complex but very important.

1.2 THE CONVENTIONAL HIGHWAY DESIGN APPROACH

The conventional highway design method considers horizontal and vertical alignments in two independent stages. Vertical alignment is established based on a predetermined horizontal one (Lee & Cheng, 2001).

The process starts by fixing several horizontal intersection points (HIPs) and then connecting these points by lines to form a horizontal piecewise linear trajectory. Horizontal curves are fitted at each HIP location of which the proposed horizontal alignment is configured. Several horizontal alignment alternatives are evaluated and the most adapted one is selected through several iterations. The vertical alignment, in almost a similar process as for the horizontal alignment, is then determined. The evaluation may take the form of cost minimisation by considering pavement, land, earthwork, user, and environmental costs for the alignment alternatives (Jong, 1998). It was shown that highway alignment optimisation in two separate stages could not yield more than a satisfactory solution (local optimum) (Jong & Schonfeld, 2003).

The method selects the final alignment solution by focusing on the detailed design elements. HIPs, deflection angles, curve radii, vertical intersection points (VIPs), tangents, grade values, and sight distances are among the design elements of highway alignment in 3D. These elements are considered necessary to depict the alignment geometrically according to the standard geometric policies described by such documents as the Design Manual for Roads and Bridges (DMRB, 1992-2008) and AASHTO design standards (AASHTO, 1994).

Thereafter, the selected design solution is made ready for implementation by defining point distances along the centreline of the alignment, called stations. Stations are point distances along the centreline of the alignment from a known start point by which the alignment segment lengths and their point positions are located and determined. These stations are considered to be the key data for setting out the design on the ground. Special procedures are used to transform these data onto the ground especially at the horizontal and vertical curve locations. The conventional setting out method is also considered to be exhausting and requires specific procedures to calculate some distances and to perform the tasks.

The conventional highway alignment development process is therefore considered inaccurate, tedious, and expensive in terms of time.

1.3 MOTIVATIONS

The traditional consideration of alignment selection, as explained in the above sections, consumes much time and effort, and the solution quality could not be more than a satisfactory solution (local optima). Researchers have tried to model highway alignment as an optimisation problem. The attempts were to reduce the total cost, speed up the process, and find better solutions (Jong, 1998). This was accomplished by developing mathematical and computer-based models using different methods and techniques. Calculus of variations by Shaw and Howard (1982), numerical analysis by Chew et al (1989), linear programming by Easa (1988), and genetic algorithms by Jong (1998), Kang (2008), Fwa et al (2002) and Tat and Tao (2003) are some of the studies that have been found.

The existing models differed from each other by the optimisation techniques used, the alignment representation, the cost components, and the design constraints. These models were almost the same in one point: the use of the same traditional (existing) design approach of highway alignment as a base for the model formulations. None of these studies has introduced a new design approach for highway alignment. Moreover, none of the studies has explicitly dealt with considering the alignment as part of a road network system through providing access points as junctions, with which the definition of optimum highway alignment solution may vary.

Since its introduction, despite the extreme development in computers and highway surveying field instrument technologies (e.g. total station), highway engineers and planners and even the researchers are still using the same design approach. No single study has exploited the technology development to specifically explore the possibility of changing some ideas imposed on highway alignment planning and design. A question arises here: would it still be necessary to keep the same planning and design approach or would it need to be changed to reflect technology development? That is one of the questions that this study seeks to answer.

The complexities in developing a model for highway alignment optimisation are due to the number of the design elements involved, their relations, and the formulations required to interpret them to a realistic highway alignment. Therefore, in order to efficiently evaluate an infinite number of alternatives in a continuous search space, such as for highway alignment, a method that simplifies the development process would still be needed (or is still demanding). Therefore, this study suggests a new design approach for highway alignment configuration. It tries to design highway alignment directly through the station points without the need to use IPs, tangents, and curve fittings.

Figure 1.1 shows the differences between the traditional and the proposed design approaches.

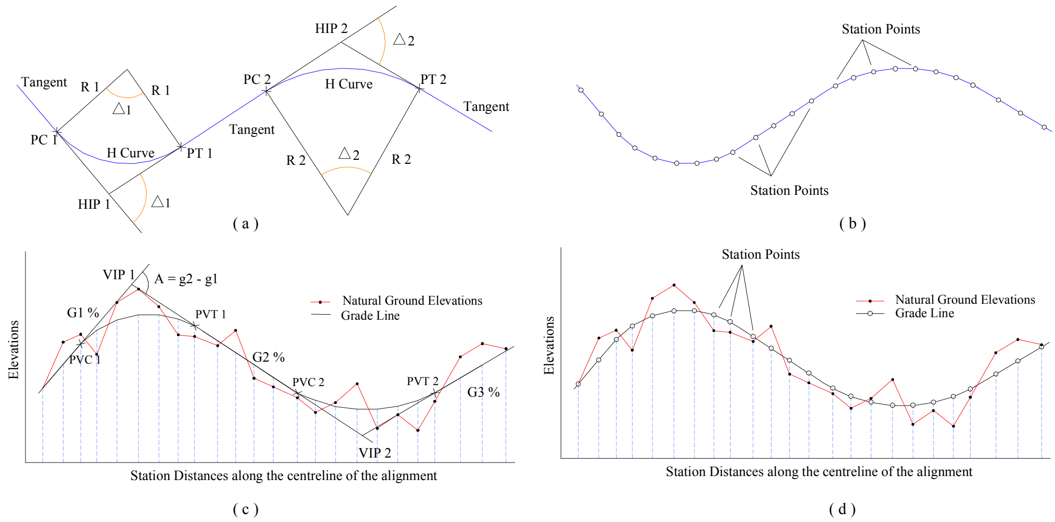


Figure 1.1: a & c) Horizontal and Vertical Alignment configuration with the traditional method, b & d) Horizontal and Vertical Alignment configuration using station points

1.4 RESEARCH AIMS AND OBJECTIVES

The highway alignment development problem is to select an economical path between two termini points and promote the sustainable development of the area under consideration through providing access to the different land uses of the study area. Thus, in recognition of the importance of the project, the overall aim of this study is to investigate the possibility of improving the highway alignment development process by the use of modern technologies and, as a sub-aim, to deal with highway alignment as part of a road network system through considering junctions for access provision.

In order to achieve the above stated aims, the study has set the following objectives:

1. Investigate the possibility of using station points along the centreline of the proposed highway alignment as a new design approach for simultaneous horizontal and vertical alignment development.
2. Ensure that any alignment produced by any model proposed as part of the research satisfies horizontal and vertical curvature directly through the station points' configuration based on predefined radius and K values, respectively. These curvature constraints are important to generate a realistic solution

according to the required design policies for safe and comfortable driving conditions. Limitations for vertical gradients are also included.

3. Investigate the possibility of incorporating access points in the form of junctions with the proposed design approach of station points, and then investigate the effect of access provision and their optimum locations on highway alignments so that the idea of simultaneous junction-highway alignment optimisation is discussed and, if successful, is introduced. This is expected to change the definition of highway alignment optimisation from a single highway alignment problem to a joint junction-highway alignment optimisation problem. The idea of simultaneous junction-highway alignment optimisation would promote a sustainable development of the area under consideration.
4. Investigate the possibility of using GA, as a search tool, for solving the problem of highway alignment development with the use of the proposed design approach.
5. Test the viability of the proposed approach for highway alignment development.

To achieve the above objectives, understanding and identifying the fundamentals of the problem components becomes essential for introducing the new design approach and to developing a highway alignment. Therefore, the following sub-objectives have also been defined:

1. To find out how the problem was dealt with in other researchers' work and identify the main components of a good model.
2. Identify the factors that affect highway alignments.
3. Understand the characteristics of highway alignment geometric design elements and their standard requirements for a realistic alignment and safe driving condition. This is important in order that all relevant characteristics can be incorporated into any approach using station points.
4. Identify the cost components that are thought critical to the problem.
5. Understand the principles of GIS and the ways that it deals with spatial and geographic data. This is thought important to find a proper way to format the study area in the suggested model. This includes developing a testing regime

(worlds and the features that would be included) which reflects the practical situation that might arise in highway alignment planning.

6. Understand the elements of sustainability and consider them for both the main alignment development and when access provisions are considered for the land use centres.
7. To investigate the possibility of formulating and solving the optimisation of a model as a genetic algorithm. In order to do this it will be necessary to investigate:
    - The possible representations of the problem in terms of the physical world and the chromosome of the problem
    - The population size for any particular problem
    - The procedures for selection
    - The procedures for crossover
    - The procedures for mutation
    - The rates for applying mutations
8. To determine whether the GA can produce optimum and realistic alignment.

It is worth mentioning that the idea of this research study is inspired by technology development and, also, it is based on the fact that any generated alignment, by whatever method, will finally consist of a series of station points and it will be implemented (set out) on the ground depending on these station points. Therefore, it is expected that the proposed approach provides immediate field data for setting out the design on the ground in the same way as providing facilities at the planning and design stages. The design data of the optimum solution, represented as the X, Y, and Z coordinates of the station points, can be uploaded to a total station, and the station points' positions are downloaded on the ground of which the exact alignment design is set out without the need for additional calculations or techniques as the traditional methods do.

If it is found that station points can be used to determine the alignment then it should be possible to combine the stages of planning, design, and generating the data required for setting out the design in one process, and consequently the development process would be simplified, and the total time and the total cost are significantly reduced.

1.5 DELIMITATION OF THE SCOPE

There are numerous factors and parameters that may require investigations for highway alignment development. This research is about testing the applicability of a new design approach for highway alignment development more than a parametric investigation; therefore, the boundary of the problem and the number of elements involved are narrowed down so that the decisions on the results are easier and clearer. The following are the main delimitations of the research:

1. Specific world sizes with specific land and topographic features are used. The features that are selected are believed to be critical in real world applications. The maximum world size dimensions that this study uses are defined as 12,000m X 6,000m with 200m x 200m grid cell sizes.
2. To represent the highway alignment, several station points are investigated. The maximum number is set as 60 station points.
3. It is required by the aim of the project to test the applicability of the proposed design approach rather than obtaining a solution with an exact (accurate) cost value. Therefore, the cost components (only those believed critical to test the new design approach) are formulated (e.g. construction cost, location cost, and earthwork cost) alongside approximation for simplification.

1.6 RESEARCH APPROACH

The research approach of the model formulation is, in general, divided into two stages of investigations: highway alignment optimisation and junction-highway alignment optimisation.

1.6.1 HIGHWAY ALIGNMENT OPTIMISATION

Highway alignment optimisation forms the main part of the study. The formulation of this part is made in steps so that the model parameters are carefully tuned and the weak and strong points of each stage are readily disclosed to avoid being confused with the components of the other stages. The study focuses on formulations for horizontal alignment optimisation (2D), simultaneous horizontal and vertical alignment optimisation (3D), and constraint handlings. These are detailed in Chapters 4, 5, and 6, respectively.

1.6.2 JUNCTION-HIGHWAY ALIGNMENT OPTIMISATION

Access provision in terms of junctions is integrated with the problem of highway alignment development. The model of section 1.6.1 above is extended so that it can also handle the problem of access provision. This is to investigate the effect of junctions and their locations on the main highway alignment optimisation. Chapter 7 gives the detail.

1.7 STRUCTURE OF THE THESIS

The structure of the thesis with a brief overview for each chapter is as follows.

Chapter 1 Introduction

Introduces the research background and motivations, defines the problem, highlights the research aims and objectives, and presents the research approach.

Chapter 2 Highway Alignment Development

This chapter reviews the literature on (i) background of the highway alignment development process, (ii) access provision and sustainable development with regard to highway alignment projects, (iii) various cost components, constraints, and highway alignment evaluations. The chapter also briefs some ideas about public participation and decision makers in the highway alignment development process.

Chapter 3 Highway Alignment Optimisation Models

This chapter reviews the literature regarding the existing models for highway alignment design. It focuses on the techniques, design principles, the cost components, the constraints, and the solution qualities.

Chapter 4 An Optimisation Model for 2D Horizontal Highway Alignment

This chapter presents developing a model for 2D highway alignments using the proposed design approach of station points. The cost and the GA formulations are discussed in detail. The chapter presents some experiments used to test the efficiency of the model.

Chapter 5 An Optimisation Model for 3D Horizontal Highway Alignment

The structure of this chapter is similar to Chapter 4. It is dedicated to formulating a model for 3D highway alignment optimisation.

Chapter 6 Constraint Handling Algorithms and the 3D Highway Alignment Model

This chapter deals with the critical constraints of highway alignment like horizontal curvature, vertical curvature, and vertical gradients, in addition to considering earthwork cut-fill balancing as constraints, too. Moreover, the chapter presents some sensitivity tests for the model against different planners', designers', and decision makers' requirement and presents experimental results.

Chapter 7 Junctions and Highway Alignment Optimisation

This chapter shows why access provision is important and investigates the possibility of building a model using the proposed design approach of station points for junction location optimisation. The chapter also investigates the effects of optimum junction locations on the main highway alignment.

Chapter 8 Conclusions and Further Work

The research contribution in relation to the stated aims and objectives of the thesis is set out in this chapter. Recommendations for future work are also summarised.

CHAPTER 2

HIGHWAY ALIGNMENT DEVELOPMENT

2.1 INTRODUCTION

This chapter presents the main principles of the existing (traditional) method for highway alignment development. It starts with defining the process and the factors that affect route location, planning, and design. This chapter also describes how providing access is important and how it affects the alignment. Sustainability elements and their consideration in highway alignment planning are also discussed. Moreover, some critical cost components to highway alignment are also presented.

2.2 HIGHWAY ALIGNMENT DEVELOPMENT PROCESS

There are usually several basic sequential stages in the highway alignment development process: planning, preliminary design, final design, right-of-way, and construction. Then, the maintenance will start and continue throughout the life of the facility (FHWA, 1997).

At the planning stage, it is necessary for the proposed facility to be defined at local and/or regional level depending on the size of the project. This process ensures public involvement and provides input into the decision-making process. Problem definition by the public and participant parties and the agreement on the proposed project are the key sources for the development to take place (FHWA, 1997).

Wright and Ashford (1998) stated that, after the need for a new route between two termini points is considered, the location of the facility (route) needs to be established among several alternatives. Following the decision on the final alignment location, the design and construction stages of the facility will start. They also added that the design and construction processes include detailed horizontal and vertical alignments, dimensions, slopes, construction standards, and quantities and type of materials.

Kalamaras et al (2000) described the process of highway alignment selection as a multi-disciplinary decision problem. They considered four principal objectives for the highway evaluation. These objectives were: minimise the impact on the environment, maximise functionality in terms of safety, minimise the construction cost, and maximise the results of the economic investment. It was stated that each criterion could be given a weighting factor to signify the importance of the component in the decision-making process.

Piantanakulchai (2005), in the same way as FHWA, described that the decision on highway planning and construction is a complex process since many objectives and stakeholders are involved. However, he added that an alignment selection may differ among different practitioners and stakeholders through the weights being given to the alignment's components. Piantanakulchai explained that the highway development and alignment selection process, as agreed by Kalamaras too, is a multi-disciplinary decision problem, clarifying that the objectives of a group for an alignment may conflict with one another if their views differed. To evaluate alignment alternatives, Piantanakulchai stated that cost/benefit analysis based on tangible and intangible factors needs to be carried out reasoning that highway construction stimulates economic generation of the region and has long-term effects on the community.

2.3 THE MAIN CONCEPTS OF ROUTE PLANNING

The planners and designers have given different emphases to the factors that affect route planning and design. For this purpose these approaches in the literature were reviewed so that the factors that affect route planning are identified. The following describes the different attitudes in route planning.

A plan for a route is a complex process and the most important part to highway alignment development is the location. Corridor or location selection of a highway not only affects the cost and operational efficiency of the facility but also influences the nearby communities and the environment. Therefore, in recognition of the importance of this task and during location planning of a route, considerations of technical, economical, social, environmental, and ecological factors will be of great interest (Wright et al, 1998).

On the other hand, Wright et al (1998) consider that the corridor location selection is an iterative procedure because, before the decision is made on a final route location, a preliminary location needs to be determined, designed, and then evaluated with regard to cost, social, and environment requirements. They described that, if the initially selected location does not satisfy the objectives, other locations are examined and evaluated until almost arriving at a global balanced location with respect to the involved interests.

O'Flaherty (1991) describes an approach for highway alignment planning as a "hierarchically structured decision process". He subdivided the process into several sequential steps. The first step is to fix the termini points and define a region (region A in Figure 2.1) of about one-third wide of its length. Then, narrowing the region by searching in a number of broad bands (bands B and C in Figure 2.1) within which the corridors D, E, and F could be selected. Then, an alignment is generated within the most suitable corridor. This approach, as described by O'Flaherty, is a continuous searching and selecting process which increasingly requires more detailed knowledge at each decision-making stage. It was also stated that, other than "tangible" factors like topography, soil and geological details, land use, population distribution, travel demand, user costs, structure and maintenance costs, and safety, "intangible" factors like political, social, and environmental issues need also to be considered with the public consultation for the alignment selection process before the final decisions can be made.

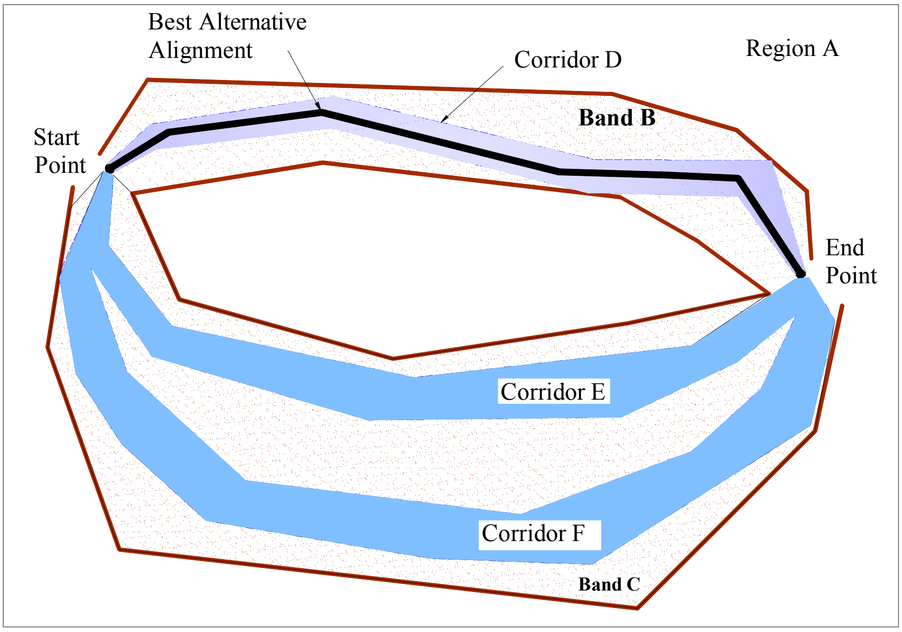


Figure 2.1: Stages of alignment selection

O'Flaherty (1991) stated that the decision on a highway location is complex and the factors that affect the process are interrelated. These factors were defined as engineering, social, ecology, and aesthetics such that the route selection process has to result in a lowest construction, land, traffic, and environmental costs.

So far, it can be seen that both Wright and O'flaherty have common views that the highway alignment location determination is a complex and repetitive process due to the involvement of many conflicting factors. They implicitly agreed that a highway location should also enhance social and economic activities which cannot be satisfied unless access is provided to the different land uses.

Finding the optimum path between two termini, as stated by Parker (1977), is a function of one or a number of physical, social, economic, design, aesthetic, and environmental factors and the best solution needs to be selected among an infinite number of possible combinations of lines and grades.

Jong et al (1999) referred to topography, soil conditions, socioeconomic factors, and environmental impacts such as air pollution and noise as the main factors in optimal path selection for an alignment.

Zhang and Armstrong (2008) as well as Maji and Jha (2009) considered corridor location as a multi-objective problem. Zhang stated that the participation of multiple stakeholders with different interests and emphases (e.g. engineering and environmental) are required to establish a good plan for a corridor location and it will be difficult to locate a single corridor that is best for all. The optimality, according to Zhang, will often have different interpretations among the decision makers. From this viewpoint, it can be concluded that the purpose of the road may determine the component(s) that should be given more weight.

Garber and Hoel (2009) stated that location determination is an important initial step in highway alignment design, and they agreed on the same elements as described by Jong et al (1999). They paid a high attention to earthwork cost due to its effect on the economic evaluation of the highway and the decision on the final location. They added that considerations must also be given to environmental impact, reasoning that any highway should cause minimum disruption to archaeological sites and to the other land use activities. This would determine that the location cost of the environmentally sensitive areas should be priceless to encourage the alignment location during the alternative evaluations to avoid such valued areas.

Saffarzadeh et al (2007) agreed that the best corridor selection among a variety of possible alternatives is a complex, tedious (repetitive), and time-consuming process and, therefore, the traditional design procedures have failed to accurately locate the best position. They mentioned that mathematical and computerised models have been developed to enable the planners and the design engineers to speed up the process.

It is well worth mentioning that the first computer models for establishing route locations were developed as early as the 1960s by Roberts and Suhrbier (1966; as it is cited in Easa et al, 2002).

Easa et al (2002) believe that many factors affect the general form of highway alignment and tend to bend the path from a straight line to a winding one. They listed the most influencing factors as topography and terrain elevations (e.g. mountains and valleys), environmental features (e.g. rivers, lakes, and wetlands), and man-made and land use objects (e.g. residential areas). He stated that all these elements affect the overall alignment cost and greatly influence the form and location of the route. Easa et al overlooked the fact that an alignment would also tend to bend when providing access to an important land use activity; this becomes inevitable while keeping in consideration that the minimum distance between the alignment location and the land use is preserved within the defined limit.

In this regard, Saffarzadeh et al (2007) stated that the minimisation of earthwork is the major factor of alignment optimisation in mountainous areas, while the length and the environmental features where the alignment passes are the main factors in level and rolling terrains. Saffarzadeh et al also pointed out that the alignment, in some circumstances, should fulfil some restrictions like compulsory points that the alignment should pass (e.g. areas of recreational interest) or points that the alignment should avoid (e.g. military and political areas). These factors affect highway corridor location regardless of the total costs. A similar concept was adopted by Cheng and Lee (2006) in their model for highway alignment optimisation, of which some details are given in the next chapter.

The above literature approaches agree that the highway alignment development process is affected by social, environmental, and economical factors, and while planning for a road location, special considerations should be given to land features' costs, topography, construction costs, traffic, and soil conditions. These approaches, although implicitly describing that the environmental issues and air pollution have impacts on the communities, they do not explicitly explain how these affect the route and its location. These communities require efficient access to their destinations so that the socio-economic generation of the whole area is promoted. This issue needs to be clearly addressed when a highway project is proposed for an area.

The effect of these factors on the route location to gain the acceptance of the public and/or decision makers varies among different project purposes. The purpose of the highway to be built and the interests of the stakeholders may differentiate among these conflicting criteria. However, a solution in order to be globally optimal should conform to the design standard criteria and compromises among the different interests.

The alignment development should therefore consider land costs, construction costs, and operational costs while meeting the requirements of technical constraints, preserving the environmental areas, and enhancing socio-economic activities. These will satisfy the decision makers by reducing the costs and satisfying the public through the development of the area on a sustainable basis.

2.4 HIGHWAY ALIGNMENT AND GEOMETRIC DESIGN

Finding the dimensions of the geometric design elements of highway alignments is called geometric design. Horizontal and vertical alignments are two major components of highway geometric design. The essential consideration of geometric design, as stated by Garber and Hoel (2009), is to provide smooth and safe highway alignments for driving to satisfy the requirements of both the drivers and the vehicles that use the road.

Each country has a discrete standard specification manual in which the limitations and standard requirements of different highway classes are defined according to the conditions of that country. Krammes and Granham (1995), by giving examples of 11 developed countries, showed that the similarities in the alignment design policies among these countries are more than the differences. The UK's design manual for roads and bridges (1992-2008) and the American Association of State Highway and Transportation Officials (AASHTO, 1994) are two examples of standard specifications for highway facilities in the UK and USA, respectively.

In essence, highway alignment is a three-dimensional problem represented in the X, Y, and Z coordinates (Figure 2.2 a) which make the design computation a cumbersome process (Krammes and Garnham, 1995; Mannering et al, 2009). Furthermore, the implementation and construction of a design based on three-dimensional coordinates is complex and difficult, too. Therefore, the problem is reduced to two two-dimensional alignment problems for simplicity (Goktepe et al, 2009; Hassan et al, 1998). One corresponds to the X and Y coordinates and is referred to as horizontal alignment (Figure 2.2 b) and the other corresponds to the X and Z coordinates and is referred to as vertical alignment (Figure 2.2 c). The existing design approach has made further simplification on the alignment coordination system. Distances along the centreline of the alignment are measured from a specified start point to define different point locations of the alignment. These distances along the alignment are referred to as stations (Figure 2.3) and are used to assist in setting out the alignment on the ground.

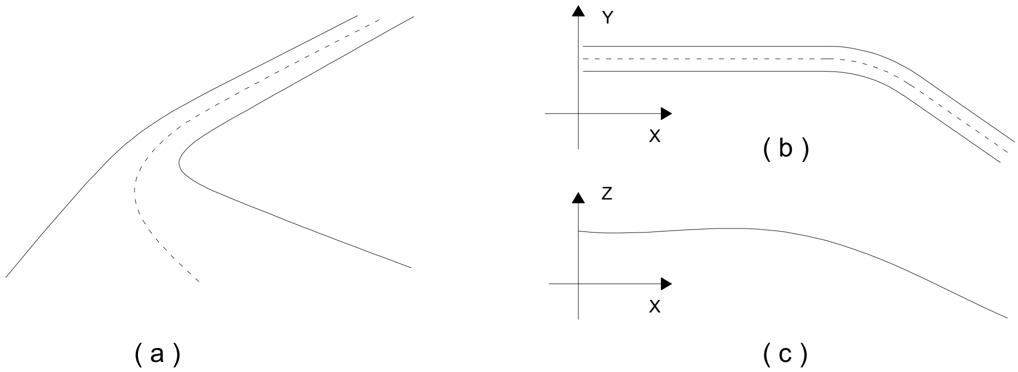


Figure 2.2: (a) A 3D highway alignment, (b) 2D horizontal alignment, and (c) 2D vertical alignment

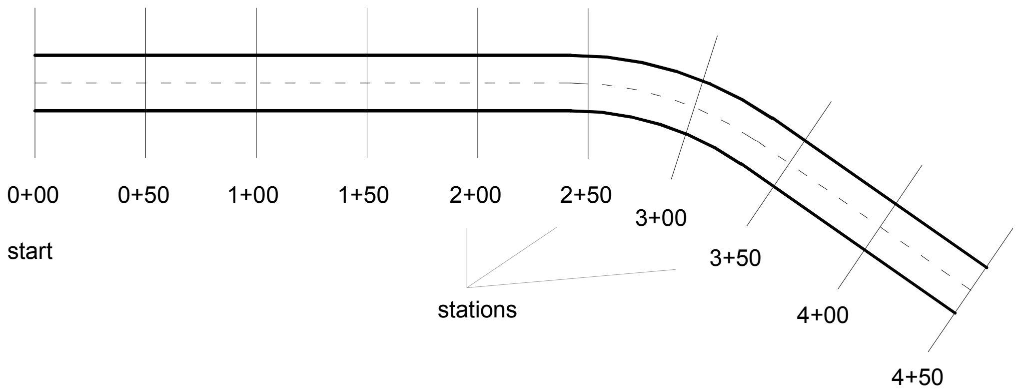


Figure 2.3: Distances along a highway alignment in term of stations

2.4.1 HORIZONTAL ALIGNMENT OF A HIGHWAY

The horizontal alignment of a highway is the plan view of the route. It consists of straight sections (tangents) of the roadway connected by curves. Curves are provided for smooth transition to avoid sudden changes in direction. Seeking a low cost location for the horizontal alignment is advisable; however, the planner and/or designer must adhere to the design standards that may produce a more expensive alternative.

In the design of a horizontal alignment many elements are involved in the calculation process. The design values of some of the geometric design elements are constrained by the standard design requirements. The elements are: Horizontal Intersection Point (HIP), Tangent distance (T), Point of Curvature (PC), Point of Tangency (PT), Length of the curve (L), Length of chord (Lc), External distance (E), Middle ordinate (M), Intersection or deflection angle (Δ), Radius of the curve (R), Stopping Sight Distance (SSD), Passing Sight Distance (PSD), and Degree of the curve (D_c). Furthermore, the design speed (V) of the highway is another influencing entity that affects most of the mentioned design elements. Figure (2.4) shows a typical layout of a simple horizontal curve.

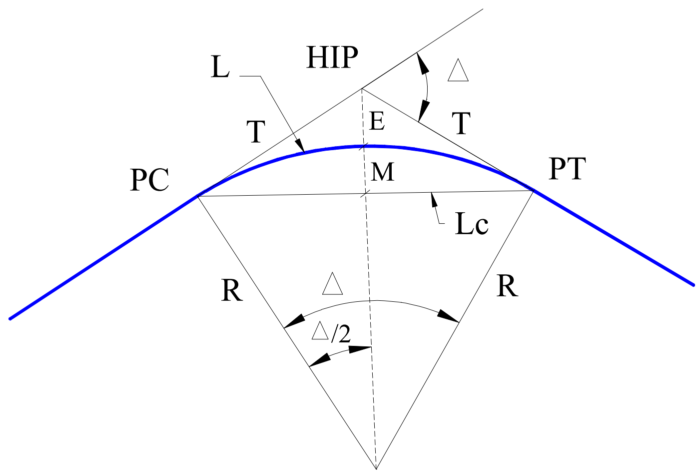


Figure 2.4: A typical layout of a horizontal curve

The design elements are required by the existing design approach to geometrically represent the alignment, simplify the calculation, and to determine the value and locations of these elements (e.g. HIP, PC, and PT). Moreover, these elements are geometrically interrelated and formulated in equations to find and, at the same time, to restrict one by each other. The determinations of these elements are used to configure the alignment and, consequently, to derive the station distances along the centreline of the alignment.

The equations and the relations that define the design elements are listed below:

$$Curve\ Radius, \qquad R = \frac{V^2}{g\ .(e + f_s)} \tag{2.1}$$

$$Tangent, \qquad T = R\ .tan\left(\frac{\Delta}{2}\right) \tag{2.2}$$

$$Length\ of\ Chord, \qquad Lc = 2\ .R\ .\sin\left(\frac{\Delta}{2}\right) \tag{2.3}$$

$$External\ Distance, \qquad E = R\ .\ (\sec\frac{\Delta}{2} - 1) \tag{2.4}$$

$$Middle\ Ordinate, \qquad M = R\ .(1 - \cos\frac{\Delta}{2}) \tag{2.5}$$

$$Length\ of\ the\ Curve, \qquad L = \frac{R\ .\Delta\ .\pi}{180} \tag{2.6}$$

Where g, e, and f_s are the gravity force, superelevation, and side friction, respectively.

Further to the design speed on the proposed road, radii of the curves have also critical influence on the movement transition and curve smoothness between the successive tangents. The transition between any two successive tangents is constrained by the design speed and the defined radius of which the standard requirements must be satisfied.

The smoothness of horizontal curves is either described by their radii (e.g. 200m curve radius) or by the degree of the curve (D_c) (e.g. a 4 degree curve). Both show the degree of smoothness or sharpness of the transition between the two tangents for safe driving.

Degree of the curve is defined using two methods, arc definition and chord definition. Arc definition defines degree of the curve as the central angle (D_c) that is subtended by a 30.5m (100ft) arc length along the horizontal curve, while chord definition considers an angle (D_c) as the central angle that is subtended by a 30.5m (100ft) chord length. Different standard specification defines different arc and chord lengths. Both definitions are presented in Figure 2.5.

It should be noted that the curvature of the horizontal alignment, which is constrained by design speed and either radius or degree of the curve, is considered very important for safe driving conditions and appearance of the alignment.

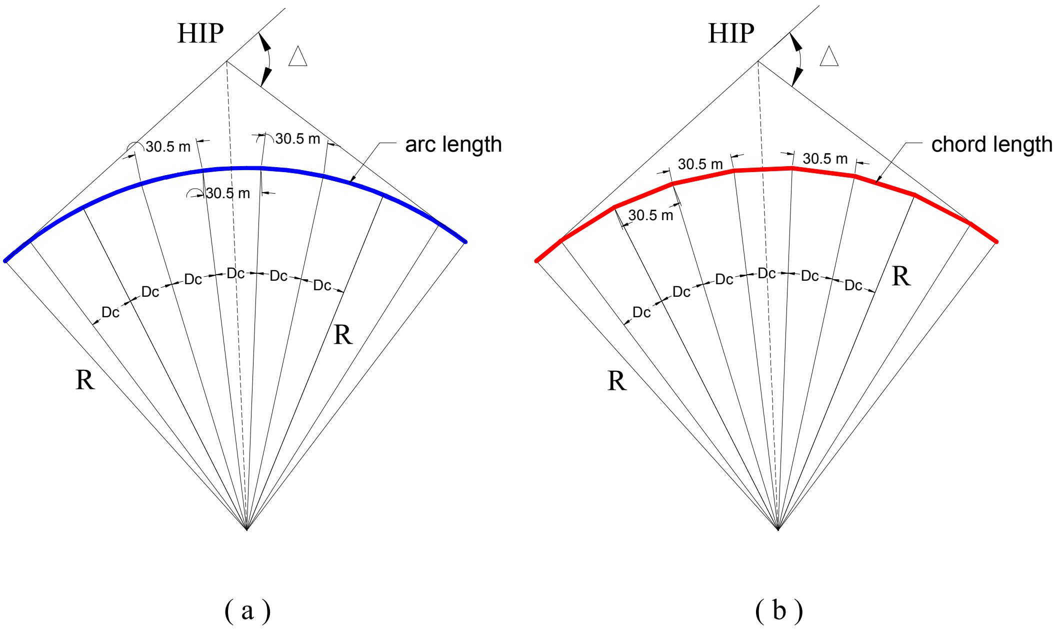


Figure 2.5: Degree of the curve, (a) arc definition (b) chord definition

The expressions for both definitions are as below:

$$R = \frac{5490}{\pi.Dc} \quad (for\ arc\ definition, using\ metric\ system) \tag{2.7}$$

$$R = \frac{15.25}{\sin(\frac{Dc}{2})} \quad (for\ chord\ definition, using\ metric\ system) \tag{2.8}$$

Using these formulas, the degree of the curve can thus be determined if the radius is known, or vice versa.

2.4.2 VERTICAL ALIGNMENT OF A HIGHWAY

The main purpose of vertical alignment is to provide point elevations along the centreline of the horizontal alignment. These elevations must be designed in such ways that satisfy user safety, user comfort, and easy drainage for rain water.

Vertical alignment is configured from connecting straight sections known as vertical tangents or grades using parabolic vertical curves. Two forms of vertical curves are available: crest vertical curve and sag vertical curve. Curves are used to provide a gradual change in elevation between successive tangents for smooth traverse of vehicles. Therefore, and based on the existing design approach which is still in use by the highway planners and designers, the selection of proper grades for the tangent sections and proper curve lengths is the aim behind vertical alignment design.

As with horizontal alignment design, vertical alignment consists of several design elements. These elements are constrained by the standard design limits and operational requirements. The vertical design elements are: Vertical Intersection Point (VIP), Grade (±G%), Algebraic difference in grade (A), Length of the vertical curve (L_v), Stopping Sight Distance (SSD), Point of Vertical Curvature (PVC), and Point of Vertical Tangent (PVT). Figure 2.6 shows a typical layout of simple crest and sag vertical curves.

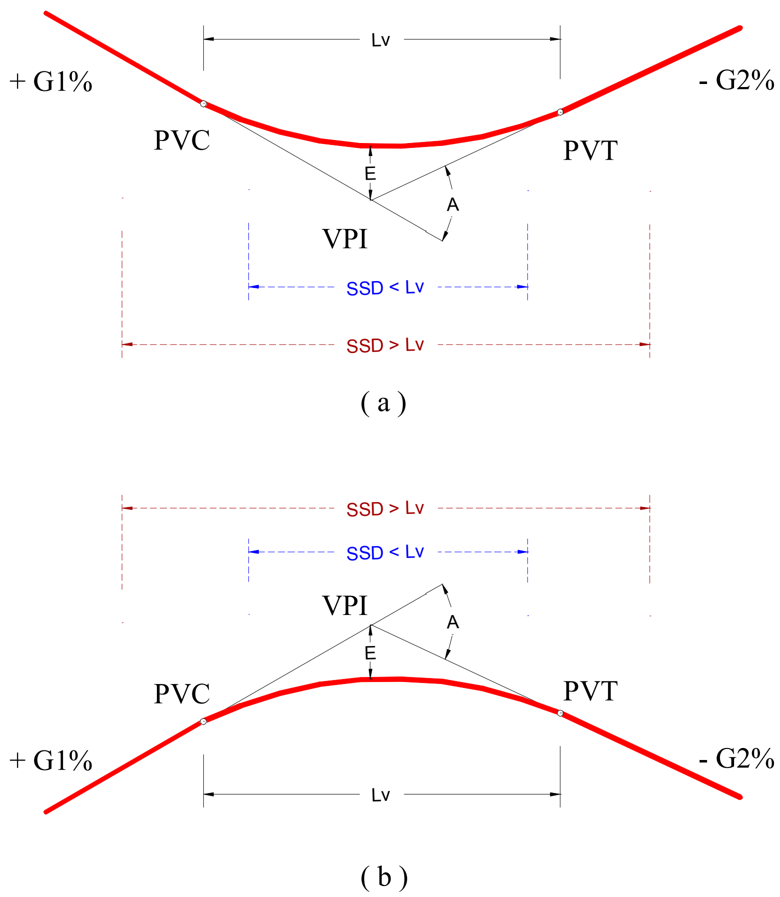


Figure 2.6: (a) Sag vertical curve, (b) Crest vertical curve

The minimum required curve length calculation is based on required SSD length. Two options are common for vertical curve length calculation; one when SSD is greater than Lv and the other when SSD is less than Lv. Both expressions for vertical curve length are calculated as follows.

On sag vertical curves:

$$Lv\ min = 2\,.SSD - \frac{120 + 3.5\,SSD}{A} \quad (m, for\ SSD > Lv) \tag{2.9.1}$$

$$Lv\ min = \frac{A\,.SSD^2}{120 + 3.5\,SSD} \quad (m, for\ SSD < Lv) \tag{2.9.2}$$

On crest vertical curves:

$$Lv\min = 2\,.SSD - \frac{658}{A} \quad (m, for\ SSD > Lv) \tag{2.10.1}$$

$$Lv\min = \frac{A\,.SSD^2}{2158} \quad (m, for\ SSD < Lv) \tag{2.10.2}$$

The design length of vertical curves can also be calculated based on the K factor. The K-value is the horizontal distance, in metres, required to effect a 1% change in the slope of the vertical curve (Mannering et al, 2009). In other words, Garber and Hoel (2009) define the K factor as the length of vertical curve per percent algebraic difference in intersecting grades (A). The expression of using the K factor for vertical curve length calculation is:

$$A = |G_2 - G_1| \tag{2.11.1}$$

$$Lv = K\,.A \tag{2.11.2}$$

This formula is used based on the assumption that SSD is less than the curve length (SSD < Lv) (Garber and Hoel, 2009; Mannering et al, 2009). Garber and Hoel do not suggest vertical curve design for SSD > Lv unless the design speed value is raised by three times. Mannering (2009) declared that some discussion about the assumption of SSD < Lv is warranted under the following justifications:

1. This assumption is almost always made in practice
2. In many cases Lv is greater than SSD
3. When SSD is greater than Lv, the assumption of Lv > SSD gives longer curve length thus producing a smoother and safer design.

K value is therefore calculated based on a known SSD, which is a function of design speed, as in the equations below.

$$K = \frac{SSD^2}{658} \ (m, for\ crest\ vertical\ curve) \quad (2.12.1)$$

$$K = \frac{SSD^2}{120 + 3.5\,SSD} \ (m, for\ sag\ vertical\ curve) \quad (2.12.2)$$

The K values of various design speeds for both crest and sag vertical curves are shown in Tables 2.1. More details of K values are given in Appendix 1.

It is worth mentioning that the new proposed design approach of station points adopts the K value for vertical curvature satisfaction. The detail is presented in Chapter 6.

Details about what has been mentioned above for highway geometric design can be found in Woods et al (1960), Ashworth (1966), Salter and Hounsell (1996), AASHTO (2004), Mannering et al (2009) and Garbe and Hoel (2009).

Table 2.1: Design values of K for Crest and Sag vertical curves based on SSD

Design speed (mi/hr)	Rate of vertical curvature for CREST vertical curves, K*		Rate of vertical curvature for SAG vertical curves, K*	
	SSD (m)	Design K	SSD (m)	Design K
15	20	1	20	3
20	35	2	35	6
25	50	4	50	9
30	65	7	65	13
35	85	11	85	18
40	105	17	105	23
45	130	26	130	30
50	160	39	160	38
55	185	52	185	45
60	220	74	220	55
65	250	95	250	63
70	285	124	285	73

*Source: AASHTO (2004), Mannering et al (2009) and Garber and Hoel (2009).

The above geometric design elements of both horizontal and vertical alignments are important to configure the alignment. They are interrelated and require complex procedures to calculate each of them so that the alignment is configured according to the standard geometric policies. One of the difficulties with the traditional design method is that, when an element design value is changed, due to any design or location limitations, the entire design elements will be affected and need to be recalculated. This may also lead to relocation and necessitate recalculating the subsequent curve locations as well. This could also be the case during the implementation of the project when the alignment location, at any part along the alignment, is reallocated. Thus, a small change would consequently affect the whole development process. Despite the technology developments of computers and highway field instruments (like total station), no changes on the alignment development process have been introduced to reduce the complexity in the development process.

2.5 FIELD SETTING OUT OF HIGHWAY ALIGNMENTS

After the alignment design has been accomplished, the design is interpreted by means of stations along the centreline of the alignment. These stations are considered the key element for the alignment setting out on the ground. Moreover, the positions of the alignment station points on the ground require specific procedures and tasks in order to facilitate the setting out process. The conventional manual procedures of setting out highway alignments are therefore considered expensive in terms of time. The methods that are commonly used for this process are deflection angle, offset from the tangents, offsets from the long chord, and coordinates alongside some other methods. The details about these methods and the way that highways are set out on the ground can be found in text books such as by Schofield and Breach (2007), Uren and Price (2010) and Allan (2007).

It can be seen that the design and setting out of highway alignment are two independent processes and the tasks require the implementation to differ from that required with the development process. Moreover, it was found that there is also an intermediate stage between the design and the implementation, which is the interpretation of the alignment geometric design by means of stations. This is unavoidable by the traditional methods, so that the alignment meets the requirements of implementation tasks. Therefore, the integration of all these tasks may help reduce the time required for the entire processes and pose simplification in order to switch between the design and implementation tasks.

2.6 JUNCTIONS AND HIGHWAY ALIGNMENT DEVELOPMENT

The presence of different settlements and different land uses in an area may enhance socio-economic activities and activates more traffic generation. Setting a system for traffic movement generates more movements between these settlements. Thus, providing access to these settlements will enhance the socio-economic activities of the area (Bruton, 1992). Kitamura and Fujii (1997) stated that, in addition to safety, energy consumption, travel time, and economic growth, a good plan for a transportation project should also satisfy accessibility.

Access points along highway alignments are referred to as junctions. Junctions are used to provide access for the nearby communities, business centres, and other land use locations. These access points might take the form of intersections (at grade intersection) or interchanges (off grade intersection). Interchanges are the most common type of access points on major highways.

Studies on junction locations and their effects on alignment location and configuration are rarely found in literature, whereas a proposal for a highway alignment should also consider the communities and the land use activities in the area where the alignment is built. This would be important to increase the serviceability of the project.

Woods et al (1960) explain that a straight line between two known termini points, if there is no specific constraints like soil and topographic limitations, is the shortest and most economical route while, by having intermediate controls, the route would tend to bend away from being a straight line. In addition to mountains, lakes, political areas, and special soil conditions, Woods stated that traffic generating centres also have significant impact on alignment location, which tends to bend the line toward these centres to increase the serviceability of the route. The route is then determined based on economic analysis of various alternatives in which traffic service costs should also be balanced (Woods et al, 1960).

Jha and Kim (2006) investigated the effect of access and proximity costs on highway alignments. Both costs were represented as accessibility costs and formulated as functions of an instantaneous distance (d) from the alignment to the centre of a property (Figure 2.7). Jha and Kim summed up the two costs to produce a balanced cost between the disutility of living close to a road and the disutility of being far from easy access. The relation of the two costs is presented in the following equation and graph (Figure 2.8):

$$C_A = C_u + C_b$$

Where C_A is the accessibility cost, C_u is the proximity cost, and C_b is the access cost.

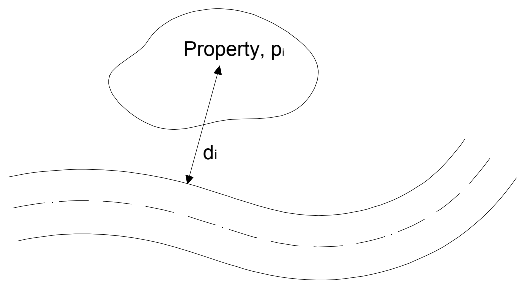


Figure 2.7: Instantaneous distance (d) between a road and a property

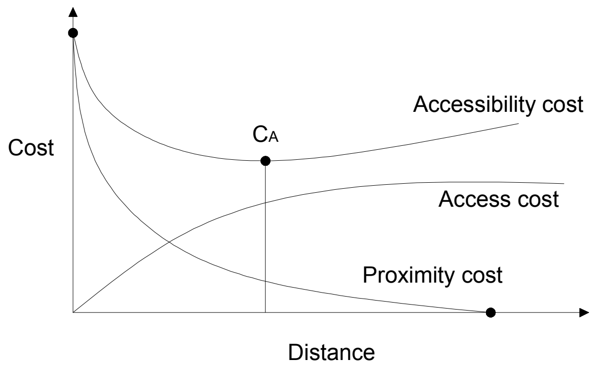


Figure 2.8: Access and proximity relationship in Jha's model (2006)

Jha's formulation is based on the fact that the access cost may locate the alignment as close as possible to the property centres while highway traffic noise and air pollution costs may push the alignment away from being located close to these properties. They attributed their considerations to the national strategic plan of the Federal Highway Administration (FHWA, 1998), which aims to protect and enhance the natural environment and communities affected by highway transportation.

Jha does not clarify in his study what part of the alignment is checked and how. In other words, Jha's model does not tell how the distance is calculated from the property centre to a point on the highway alignment, whereas the alignment only consists of tangent lines, curve segments, and point of intersections (PI). Furthermore, the study does not explain how the accessibility cost is integrated with the other objective function components. Jha's alignment is based on Jong's model (Jong, 1998), which is configured by tangents from known PI locations fitted with circular curves.

One disadvantage of this approach is that a balanced cost between access and proximity costs may not produce a practical distance between the property and the highway. For instance, a high volume of traffic may push the alignment away from the property and vice versa, while the effect of noise and pollution is likely to dismiss at a certain distance regardless of the amount of traffic. Therefore, it would be reasonable to consider the amount of traffic to determine the main route's balanced location provided that the distance between the property and the main route is constrained by a minimum user predefined proximity distance.

The total cost, location, and configuration of an alignment with access point (junctions) differ from an alignment without these points. The length and location dependent costs are greatly affected by this issue. It can be seen that the locations of these junctions also need to be optimised so as to incur minimum user cost (access cost), link construction cost, and proximity costs. Therefore, the highway development process should also be concerned about junction location optimisation in such a way that ensures efficient access from different land uses with due consideration to sustainability elements. This may increase the serviceability of the alignment and as a consequence gain the acceptance of the public and decision makers.

Figure 2.9 below shows an ideal concept of junction provision and its effect on optimum alignment location and configuration. Figure 2.9a shows an alignment where no junctions are required. This implies that a straight alignment is ideally the "best" solution where no restrictions are available. Figures 2.9b and 2.9c illustrate that the alignment tries to bend towards the traffic generation centres to increase the serviceability of the road provided that the total cost is minimised. High serviceability is achieved by providing efficient access while preserving the environmental factors. This issue is experimentally addressed in Chapter 7 of this research study.

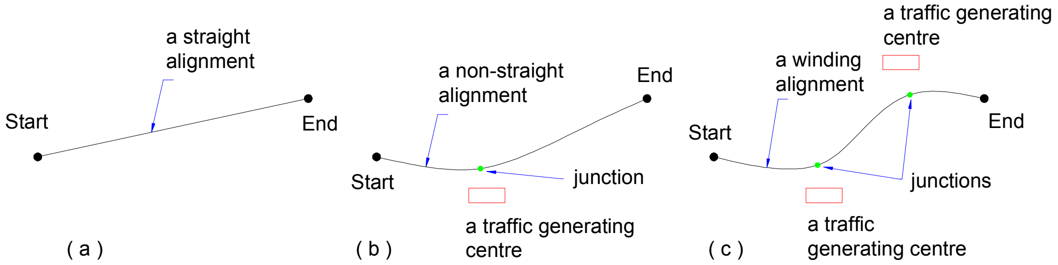


Figure 2.9: A typical illustration for the effect of access provision on highway alignment

2.7 SUSTAINABLE DEVELOPMENT AND HIGHWAY ALIGNMENT

Different definitions are available to define the concept of sustainable development. Kersten et al (1999) draws attention to the need of developing methods that emphasise the potential complementarities between economic development and environmental improvement.

In the Brundtland report (1987), sustainable development is defined as “a development that meets the needs of the present without compromising the ability of future generations to meet their own needs.”

Sustainable development, as stated by Kersten et al (1999), is affected by economical, ecological, and social elements and it tries to improve methods for natural resource handling as the demand on resources increases. These elements affect the pattern of development. Furthermore, it is stated that sustainable development seeks to produce an approach to compromise among the three elements and bring them into balance with each other (Figure 2.10).

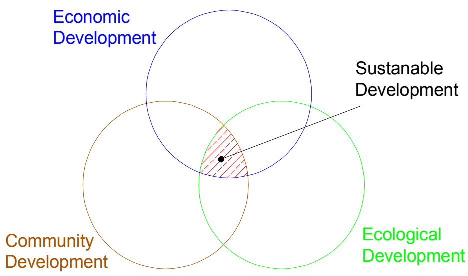


Figure 2.10: The three interrelated aspects of sustainable development

Moreover, Kersten et al show that sustainable development causes negotiations among the interested individuals and groups that are involved in the process.

Murray (2001) states that the transportation system is one of the main elements affecting the environment in terms of congestion and air pollution, emphasising that economic generations and preserving the environment are factors of sustainable urban growth. This approach can also be attributed to the disutility of a highway for being put close to urban centres and its effect on sustainable development of that area.

On the other hand, the highway alignment development process also has significant effect on land use planning. Jha et al (2006) stated that a new highway attracts new housing developments, business, and industry due to improved access, which lead to changes in land use pattern. Therefore, in order to determine a sustainable location for a highway corridor that enhances sustainable development for the whole area, careful consideration must be given to the highway alignment development process. The new strategy in transportation planning, as mentioned by Jha et al (2006), advocates noise-compatible land-use planning in such a way "that noise-sensitive land uses are either prohibited from being located adjacent to a highway, or that the developments are planned, designed, and constructed in such a way that noise impacts are minimized". Jha stresses that the planning process is to enhance and protect the natural environment through land-use control (land-use planning) and to mitigate the highway traffic noise and pollution.

In nearly the same manner, Jaarsma (1997) also agreed that land use planning of an area depends mainly on the available road network or the proposed plan for the newly suggested road network. He suggests that road network must be dealt with as a major form of land use. Jaarsma also stated that poor road network planning might lead to poor planning of other human land uses or activities associated with the road network. He described that a well-planned road network enhances economic growth, exploits land resources, and enhances social relations through the access it provides to lands.

In explaining the aims of a sustainable development, Jaarsma (1997) emphasised the need to find a balance between maximising accessibility and minimising the effect of traffic hazard, noise, and pollution as well as habitat fragmentation for the “fauna”. He adds that perspective of road planning should be broadened and the planner should move from individual road planning towards road network planning. He confirms that the aim, according to his description, is how to plan for a rural road network development that satisfies accessibility and ensures a sustainable environment.

Haq (1997) agreed that an efficient transportation system economically and environmentally is needed to move towards a sustainable development. He adds that air pollution and global warming are also affected by transportation and thus associate the process of transportation planning with the process of decision making too.

It can be concluded that the modern life depends mainly on the modes of transport, and the economic growth of an area is mainly affected by the availability of the road network and the transportation system. Therefore, during the planning stage for any highway project, the sustainability elements should be given high consideration so that the project gains the public's and decision makers' acceptance at once.

2.8 PUBLIC PARTICIPATION, DECISION MAKING, AND TRANSPORTATION PLANNING

Meyer and Miller (2001) stated that transportation planning is part of decision making, land use planning, and development processes which requires considering short and long perspectives by defining the problems, discussing alternatives and assessing future limitations. This was believed important to ensure efficient movement for goods and people.

Mayer and Miller incorporated community and public participation together as an indivisible part into the transportation planning process to facilitate satisfaction of the decision-making process. They also believed that the planning process must provide opportunities for participation to all interested groups.

On the other hand, Tang & Waters (2005) give an important role to public participation for defining the problems and discussing the possible alternatives to achieve a good transportation plan. He associated increasing the success level of planning decisions with public participation at each stage of transportation planning, reasoning that public participation at an early planning stage allows the planner to consider the requirements at the beginning of the project. Moreover, he adds that public involvement "will fill the gap between public concerns and the transportation plan developed by decision makers".

Jha (2003) explained that making a decision on a final resulting alignment corridor is awkward as it should comply with the existing environmental policies and be agreed by the interested groups and politicians because an alignment alternative would not always be selected based on its minimum costs.

It can be concluded that the consideration of public requirements should meet the satisfaction of their social, economic, and environmental needs so that the decision on the final road project plan is promoted.

2.9 EVALUATION OF HIGHWAY ALIGNMENT ALTERNATIVES

After the demand for planning a link between two points is established, the alignment alternatives must undergo the appraisal process to determine the economic, social, and environmental costs. This process enables the planners to decide on the most feasible alternative among different competing project options. In order to appraise a project, the related elements need to be formulated in a function in which the overall cost should give an indication about the suitability of the solution. Such a function is called objective or fitness function (Rogers, 2003).

Jong Schonfeld (1999) explained that many factors are involved in alignment optimisation and the cost functions should verify trade-offs among the different alignment components. He also stated that careful attention must be given to the alignment's different cost categories in such a way so as not to give a higher weight to a component on account of another one.

Saffarzadeh et al (2007) considered economical evaluation as an important factor for the decision on the best alternative and they stated that an alignment location depends on the weights that are being given to the cost components that the objective function is made of. If a high weight is given to a cost component, the search process will then give a higher chance to the associated criterion to be verified.

Economical analysis is required to compare different alternative solutions as it has a considerable influence on the final physical location of the highway (O'Flaherty, 1991).

Examples of the most known highway alignment cost components that have impact on highway alignment development are given in the following sections.

2.10 COST COMPONENTS FOR HIGHWAY ALIGNMENTS

The form of cost function formulation affects the optimal solution. Giving weight to individual elements of the problem has critical effects on the form and location of the corridor being found. Giving different weights to the correspondent elements results in different locations (Saffarzadeh et al, 2007).

As the process is complex, the traditional manual design methods find it convenient to break the alignment down into different separate components and to find a way to express all the variables in terms of a common unit (e.g. cost value) (Roberts, 1957; Chew et al, 1989; Jong, 1998). For example, the costs that mainly affect horizontal alignment may be land acquisition and environmental features. Meanwhile the most important factor in vertical alignment configuration is the earthwork cost. The horizontal and vertical alignments are interdependent and the most cost components are highly linked to both horizontal and vertical alignments; therefore, a comprehensive cost formulation for simultaneously optimising three-dimensional alignments will be needed (Jong et al, 1999).

Traditionally, costs were the most dominant criteria in highway economic analysis and are still used as a major component for highway alignment optimisations.

According to Winfrey, the Organization of Economic Cooperation and Development (OECD) in Paris, as cited in Jong et al (1999), the major cost components of highway transportation are categorised into several features, as summarised in Table 2.2 and almost similar criteria were introduced by Wright et al (1998), as shown in Table 2.3.

Table 2.2: Classification of highway transportation costs

Classifications		Examples
Planning, Design, and Administrative Costs		Consulting and supervision costs
Construction Costs		Earthworks, Pavement, Right-of-way
Operation and Maintenance Costs		Pavement, Mowing, Lighting
User Costs	Vehicle Operating Costs	Fuel, Tire wear, Depreciation of vehicles
	Travel Time Costs	Vehicle hours times unit value of time
	Accident Costs	Estimated accident rates times unit accident cost
Social and Environmental Costs		Noise, Air pollution, Wetland loss

Table 2.3: Examples of criteria in road location decisions

Criteria	Influencing Factors
Construction cost	Functional classification/design type; topography and soil conditions; current land use
User costs	Traffic volume; facility design features (e.g. gradient intersections); operating conditions (e.g. speeds, traffic control systems)
Environmental impacts	Proximity to sensitive areas; design features to mitigate impacts
Social impacts	Isolation or division of neighbourhoods; aesthetics of design; fostering of desired development patterns
Acceptance by various interest groups	Government agencies; private associations; neighbourhood groups and the general public

On the other hand, Chew et al (1989) presented the major cost components according to the percentage of each component to the total objective cost, as follows:

1. Construction costs: Drainage 10%, Earthworks 25%, Pavement 30%, Land 5-10%, Others 30%.
2. Vehicle operating costs (net present values discounted over 30 years) 30%-1000% of construction costs.
3. Design and administrative costs about 5% of construction costs.
4. Maintenance costs (net present values discounted over 30 years) about 5% of construction costs.

Jong and Schonfeld (1999), Jha and Schonfeld (2000), Jong et al (2000), and Tat and Tao, (2003) are also examples of some studies in which they classified the alignment sensitive costs, as follows:

1 Supplier Costs:

   1.1 Length dependent costs (construction and maintenance costs)
   1.2 Location dependent costs (right-of-way and environmental costs)
   1.3 Earthwork cost (cut, fill, and mass-haul costs)

2 User costs:

   2.1 Vehicle operating cost (fuel consumption, vehicle wear and tear costs)
   2.2 Travel time cost (vehicles hours x unit value of time)
   2.3 Accident cost (estimated accident rates x unit accident cost)

In addition to supplier and user costs, Jong and Schonfeld (1999) referred to the possibility of quantifying the impact of noise and pollution as well as summing up their effects with the other cost components.

Roberts (1957) and Woods et al (1960) stated that administrative and headquarters expenses do not noticeably affect the selection of location, design, and the characteristics of the proposed highway facility; therefore, their effects are neglected in economic analysis. This concept was also agreed by Jong and Schonfeld (1999), reasoning that these costs are relatively similar for various alignment alternatives of the same project whereas some other components produce different costs. Therefore, the similar cost components of different alternative alignments, in order to simplify the formulation, are usually disregarded in highway economic analysis.

However, this approach cannot always be considered as true. Different alternative alignment locations of the same project may include different structures like bridges, culverts, intersections, interchanges, tunnels etc., and require different plans, designs, and even project times which consequently impose different administration costs. Chew et al (1989) presented, as mentioned in the above sections, that the administration cost could constitute 5% of total construction cost of the road. Thus, it will be wiser to neglect its effect throughout due to its small amount rather than defining it as similar costs for different alternative evaluations.

On the other hand, it can be found that the most critical costs for highway alignment development are construction costs (pavement and earthwork costs), location costs (cost of land acquisition, special soil requirements, and environmentally sensitive areas), and user costs. Maintenance and other costs can also be included if accurate cost value is required.

2.10.1 HIGHWAY COST BREAKDOWN

The following cost components are the most common components that have been considered in literatures for highway alignment planning and design.

2.10.1.1 Location Dependent Costs

This is defined by the areas of land required for the highway right of way. The costs, such as land acquisition, soil stabilisation and high-cost environmental features like flood plains and wetlands, fall within location-dependent costs. This cost is calculated by calculating the area covered by the alignment on a specific land parcel multiplied by the land's unit cost (Chew et al, 1989; Jha et al, 2006; Jong and Schonfeld, 2003; Tat and Tao, 2003).

2.10.1.2 Length Dependent Costs

Any cost that varies with the length of the alignment can be considered within length-dependent costs. The unit construction (pavement) and maintenance costs are calculated and multiplied by the total alignment length (Jha et al, 2006; Jong and Schonfeld, 2003; Tat and Tao, 2003). The costs of construction (pavement costs) are also affected by the soil condition, as it varies from one location to another (Chew et al, 1989).

2.10.1.3 Earthwork Cost

Parker (1977) and Sthapit and Mori (1994) stated that earthwork constitutes a substantial component of the total cost for any given location project; a rural highway problem as an example. Parker (1977) stressed that, due to the extreme variations in the earthwork quantities associated with different locations, a highway location cannot be chosen independently of vertical alignments.

This cost is related to the design elevations of the vertical alignment with respect to the ground elevations. At each station distance along the horizontal alignment, the cut and/or fill cross-section areas are calculated from the difference between the road profile and natural ground elevations. Afterwards, the earthwork volume quantities are determined from these areas by applying either Prismoidal or Average-End-Area methods (Tat and Tao, 2003). The earthwork costs are then determined by multiplying the earthwork quantities by the unit cost of cut and/or fill. Chew et al (1989) stated that these costs (cut and fill) are also functions of mass haul costs and are affected by the cutting depths and soil type. Some studies emphasise the need to balance the amount of cut and fill as an issue to reduce the amount of earthwork (Goktepe and Lav, 2003). However, this might not always be the case as the planning policies of the different countries may define the way they deal with land in various other ways.

2.10.1.4 User Costs

Jong and Schonfeld (1999) and Jha et al (2006) stated that the computation of user costs is less direct because the relations between user costs and highway design features are not explicit. They considered that the most dominant costs associated with the road user include vehicle operating costs, the value of travel time, and accident costs. They suggested that only those components that depend on travel mileage are included for vehicle operating cost in highway economic analysis. Meanwhile, the travel time cost is calculated by multiplying total vehicle-hours as a function of total alignment length, average running speed, and projected travel demand by the unit value of time. They believed that fuel consumption is the most significant cost of vehicle operation along with some other less important elements like tire and vehicle depreciation.

Accident cost is also one of the influential elements but, due to the complexity of causes and lack of empirical data for a non-existing highway, it will be difficult to estimate its effect on the total alignment cost. Some studies like that performed by Zegeer et al (1992) have represented accident cost as a function of length of curve, curvature, and traffic volume. The total accident cost is then computed by multiplying accident rate with unit accident cost.

On the other hand, this cost component is affected by many factors and for a precise calculation it may require an independent research study so that the implications of each factor on highway alignment development process are determined. Following are the factors that may affect user costs and require extensive investigations.

1. The existing traffic volume and the future traffic forecast. The percentage of traffic growth needs to be available to predict the amount of future traffic.
2. Directional traffic volume. This affects the calculation of the average road grade to determine the average running speed in each direction. Fuel consumption is a function of average running speed.
3. Design life of the proposed road.
4. Length of the proposed road.
5. Design speed on the road.
6. Horizontal and vertical curvatures along the horizontal and vertical alignments (smoothness or straightness of the alignments). This affects the design speed and accident rate on the road.
7. Grades of the vertical alignment and their lengths.
8. Types of vehicle (traffic components), their speed, and their percentages on the proposed road (e.g. passenger cars, trucks, and trailers).
9. Type of fuel per vehicle type (petrol and diesel), their unit prices, and unit price change during the whole design life of the proposed road.

10. Travel price per traveller's purpose. The price of the same journey for the same distance may vary among different travellers according to their annual income and the purpose of the travel.
11. Vehicle depreciation costs.
12. Accident rate and costs
13. The road type (road class: e.g. urban, rural, two-lane, multi-lane etc.).
14. Geometric features of the proposed highway.
15. Average number of passengers per vehicle travelled.
16. Power and size of the vehicle's engine.
17. Annual net profit that the alignment generates.
18. Annual traffic growth rate.
19. Annual fuel price increase rate.
20. Annual profit increase rate.

2.10.1.5 Social and Environmental Costs (access provision costs)

Building a new road incurs many development changes on the adjacent areas as well as causing environmental impacts such as air pollution, water pollution, and noise. Public awareness of the effect of transportation projects on their environment has increased and therefore in some situations more weight is given to social and environmental cost components during objective function formulation (Jong and Schonfeld, 1999).

It should be mentioned that access problem is one of the issues that has not been explicitly dealt with in the literature. However, its effect on alignment development is critical.

2.11 CONSTRAINTS AND HIGHWAY ALIGNMENT DEVELOPMENT

In general, two constraint categories can be identified in the literature: sustainability constraints and geometric design constraints. Each of these has several elements that need to be considered during the development processes.

Cheng and Lee (2006) stated that highway alignment constraints are determined either by design codes or by engineering practices. They added that some constraints must be strictly obeyed (hard constraints) while some others need observation with flexible consideration (soft constraints).

The main sustainability elements are, as previously explained, social, economic, and environmental factors. These elements mainly affect the physical location of the route. The geometric design constraints, as with sustainability constraints, also have critical impacts on the overall alignment cost, location, and configuration. The constraint criteria for each element are generally defined by the planners, public, interested individuals and groups, and decision makers' desire. Some of these constraints are defined by the design policies as the Design Manual for Roads and Bridges (1992-2008) and/or AASHTO (1994).

Saffarzadeh et al (2007) defined the sensitive environmental areas as no-go areas where the alignment must not pass. He imposed a high penalty cost on the alignment when it passes through these areas. Using this technique the entire region's sensitive areas need to be defined; practically, this would be an exhausting and complex process, especially with large worlds. This approach, in some way, is similar to the work performed by Kang et al (2007), in which they defined feasible gates for the alignment locations where several alternatives are evaluated. 'Feasible gate' is a term which is used by Kang et al (2007) to represent a narrow corridor within which a highway alignment is optimised. This approach restricts and narrows down the search under the justification of computation time resource saving. Moreover, this approach makes the entire study area discontinuous, thus ignoring hundreds of alternative solutions for evaluation and comparison. This approach might be suitable for a very specific problem where the alignment corridor is previously determined.

For proximity aspects, Saffarzadeh et al (2007) defined a minimum distance as a constraint between a particular land use and the alignment to restrict the route location. With this approach, the alignment distance to a specific property is checked with the predefined minimum permitted distance. If the distance is smaller than the minimum permitted distance, then the alignment is penalised to reduce the viability of it as a possible solution to the problem. A similar approach was adapted by Cheng and Lee (2006) when they restricted the alignment location to both the areas of interest (control points) and restricted areas (no-go areas). It is well worth mentioning that this study has adopted a similar approach for proximity consideration (see Chapter 7).

The most critical geometric design element of a highway is the radius of which the alignment curvature is defined. Alignment radius is a function of design speed, superelevation, and coefficient of side friction. Moreover, sight distance and gradient change are factors affecting the minimum length of vertical curve (ASHTO, 1994; Jong and Schonfeld, 2003). Jong and Schonfeld (2003) in his study considered maximum gradient, minimum radius and minimum length of vertical curve as the most significant constraint elements in highway alignment optimisation.

Goktepe and Lav (2003) stated that economic considerations are also important as other design controls. It was mentioned that the construction cost can be significantly decreased when cut and fill are balanced, along with considering minimising earthwork. The reduction of earthwork amount could not be achieved by setting the grade line as closely as possible to the ground line, as stated by Goktepe and Lav, but also by considering cut and fill balance.

2.12 SUMMARY

In this chapter, the main factors that affect the highway alignment development process were explained and the most influencing objective function components were discussed. Sustainability and geometric design limitations were also considered. Moreover, different viewpoints and approaches of the various literature were presented.

In the next chapter, the techniques, elements, and criteria that have been used by researchers for highway alignment optimisation models are presented and critically reviewed.

CHAPTER 3

HIGHWAY ALIGNMENT OPTIMISATION MODELS

3.1 INTRODUCTION

The aim of this chapter is to review the models that have been developed for highway alignment optimisation. The purpose was to find the main components necessary to build a 'good' optimisation model. It was set that the review should deal with the techniques, the representation methods, the study format, the cost components, and the constraints as the key elements for an efficient optimisation model.

This chapter starts with a general introductory to highway alignment optimisation. The different attempts for vertical, horizontal, and three-dimensional (3D) highway alignment such as Dynamic Programming, Linear Programming, Calculus of Variations, Numerical Search, Network Optimisation, Neighbourhood Search, and genetic algorithms (GAs) are briefed. It should be mentioned that only one GA-based model for highway alignment optimisation was found in the literature. Few others extended the approach for modifications. The review of the existing GA-based model is presented in some details in Section 3.2.3.4. The chapter also briefs the conclusions, and outlines the formulations of the main components of this research study.

3.2 HIGHWAY ALIGNMENT OPTIMISATION MODELS

The aim behind highway alignment optimisation is to develop a mathematical search model to find a global or near global optimal solution based on total objective cost minimisation (or maximise benefit) while satisfying both the sustainability and geometric design constraints. According to Kang (2008), highway alignment optimisation is a complex non-linear optimisation problem with a noisy and non-differentiable objective function. Lee et al (2009) attribute the complexity of the problem to the nature of the design elements and the requirements set by the design codes. Under the time and budget constraints and due to complex computational work, as mentioned by Lee et al

(2009), the solution by the traditional manual method is produced from only a handful of candidate evaluations. In the same manner Easa et al (2002) mentioned that the process of 3D highway alignment design requires cumbersome computations.

Goktepe et al (2009) stated that three-dimensional alignment is subjected to a large amount of design variables, several non-linear constraint equations, and numerous solution alternatives and therefore it is considered as a complex optimisation problem.

Different researchers developed different techniques to either optimise horizontal alignment or vertical alignment or both in two different stages. Fewer attempts have been made to optimise a 3D highway alignment due to the complexity of the problem. In highway alignment projects, as described by Schoon (2001), a vertical alignment is determined based on a pre-fixed horizontal alignment. Chew et al (1989) mentioned that any highway alignment solution cannot be global unless both horizontal and vertical alignments are optimised simultaneously.

In this chapter, the existing models of highway alignment are divided into three groups. These groups are described as models for vertical alignments, models for horizontal alignment, and models for 3D highway alignments. The revisions of these studies are as in the following sections.

3.2.1 VERTICAL ALIGNMENT MODELS

3.2.1.1 Dynamic Programming (DP)

Dynamic programming is a multistage constrained optimisation technique for solving decision problems, in which the problem needs to be subdivided into N number of subsequent decision stages to find an optimal strategy. At each stage, a single decision variable is iteratively searched and its optimum value is determined (in the vertical alignment problem this single variable may refer to the grade of a line segment or elevation at a specific distance or station). The output of each stage is then used as an input to the next succeeding stage. The cumulative cost calculation of the successive stages is the optimum minimum cost of the problem and the optimal decision values of each stage are the optimum solution of the problem. The overall optimisation is obtained when the last stage is solved. This approach has flexibility for either forward (from start to end points) or backward (from end towards start points) calculations (Goktepe et al, 2005; Goktepe et al, 2009; Smith, 1991)). It should be mentioned that this approach is only applicable when the total fitness function is the cumulative of the fitness values of the successive stages of the problem (Beasley et al, 1993).

Goktepe et al (2005) used this technique for the vertical alignment optimisation problem. It was stated that vertical alignment optimisation is a constrained non-linear optimisation problem which can be solved by a multistage optimisation technique. The problem was considered as a two-dimensional problem based on a predetermined horizontal alignment. This was done, as mentioned by the authors, due to the complex behaviour of the problem as a three-dimensional problem.

Goktepe's approach was based on earthwork optimisation. The problem was partitioned into grid cells (elements) by means of horizontal and vertical grid lines. According to Goktepe, the size of the grid cells affects the precision and computation efficiency of the model. Each vertical interval was then considered as a stage of the multistage optimisation problem, which through an iterative search process, an optimal elevation value of the current stage is determined. A two-dimensional matrix, along the X-coordinate distances of the alignment, was used for the ground data elevations. At each vertical grid line and based on the existing ground data and the input information from the previous stage, the alignment elevation (optimal value) was searched and determined. When the algorithm reached the last stage of the search, a set of decision variables representing the optimum solution was determined.

It should be mentioned that the same study was extended by the author (Goktepe et al, 2009) in such a way that the earthwork minimisation was based on balancing cut and fill quantities at the subsequent alignment cross-sections. The study stated that, finding the alignment's grade line as close as possible to a pre-defined centreline elevation, the elevations with which the cut and fill are balanced are likely to minimise the earthwork cost and consequently the total alignment objective cost.

Considerations were given to the maximum grade while no requirements for curvature were presented. Earthwork cost was the only component that was used in the objective function, whereas construction and other length-dependent costs also have impacts on vertical alignment optimisation. Furthermore, the model was based on a predetermined horizontal alignment and therefore the obtained alignment cannot be more than a suboptimal solution.

It is also worth mentioning that the DP approach uses a single solution, and the variables at the subsequent stages are considered independent to each other and it is difficult to inter-relate a variable value at a late stage with one at an early stage, thus losing the possibility of other combinations that may incur less cost.

For the same purpose, Fwa (1989) used DP for vertical alignment problems with due consideration to earthwork, land, materials, and vehicle operating costs subject to gradient and curvature constraints. Gradient change value was used as a function for curvature evaluation. The profile was represented by means of straight line pieces at equal horizontal distances along the centreline of a predetermined horizontal alignment. The output had no smooth transitions and therefore the alignment was not realistic. The only difference with Goktepe’s model was that the objective function used by Fwa was more comprehensive.

3.2.1.2 Linear Programming (LP)

A problem with a linear objective function can be solved by LP for either maximum or minimum objective value subject to several linear constraints. Using this method, a solution can be obtained in a finite number of steps in which a set of ‘best’ variable values are determined (Walsh, 1985).

Easa (1988) used the technique to select an optimum roadway grade combination that produces minimum earthwork costs. The model, at each step, employed enumeration to list the whole feasible grades and used linear programming to evaluate each alternative grade line that incurs less earthwork cost. At the vertical intersection points (VIPs), vertical curves were integrated but it was noticed that no consideration was given to the alignment requirement for curvature and sight distance.

Following Easa's model, Moreb (1996) used the same technique for the same type of problem. The only difference between the two models was that Easa's model considered the grade line establishments and earthwork allocations in two different stages, while Moreb's model combined the two stages in one. The alignment was divided into several small sections and the search was to select the optimum grade line elevations. The alignment sections were defined by their X (the centre of the section) and Y (the average height of the ground at this section) coordinates and the vertical alignment grade line would connect the start and end points passing through these points. The search employed an iterative procedure to evaluate different grade line values until a combination that minimises the earthwork costs was found.

It is worth mentioning that these models were tested over a limited number of intersection points (only three intersection points were used) and no realistic alignments were generated. Moreover, Moreb's model produces a piecewise linear trajectory alignment with no curve inclusion.

3.2.1.3 Genetic Algorithms

A genetic algorithm is an evolutionary adoptive search technique used to solve relatively complex optimisation problems. A brief review about the technique is given in Section 3.2.3.4 (A).

A genetic algorithm-based model considering grade length, fixed elevation points, and non-overlapping of horizontal and vertical curves was developed by Fwa et al (2002) to analyse optimal vertical alignment for highway design. The model also paid attention to maximum grade as constraints.

Fwa represented the solution space by means of vertical and horizontal grid lines. The vertical grid lines were equally distributed along a predetermined horizontal alignment length. Positions (elevations) on the vertical grid lines were regarded as decision variables. An objective function was employed in such a way that a comprised earthwork and pavement construction costs are minimised.

The solution (chromosome) in the GA structure was represented as a string of cells. Each cell matched a vertical grid line position and held an elevation at that grid line. This approach ensures that the application of the algorithm is always limited to prefixed horizontal alignments with known lengths. The approach cannot be extended for simultaneous horizontal and vertical alignment optimisation where the length of the horizontal alignment is unknown. Therefore, this representation could be regarded as a technique to solve specific problems only.

Fwa confined the search of the initial population generation by defining a narrow feasible region to avoid generating unfeasible solutions. The feasible boundary of the search space was defined by four lines, with grades (slopes) equal to the defined maximum grade value, drawn upward and downward (positive and negative directions) from the two termini points. This approach could reduce the number of unfeasible solution generations by avoiding making a search outside the defined boundary area. It was also believed that the approach could reduce the computation time of the process because of evaluating fewer unfeasible alternatives.

Another approach called the “search angle method” was also adopted by Fwa to confine the search in the consequent generations. By this method, the search was limited as a vertical space produced from the maximum positive and negative grade lines drawn from a selected point elevation on the vertical grid line located just before the position of that vertical grid line upon which the search was focused. The extra calculation of this approach may burden the process but in return the total process time may be reduced due to generating fewer unfeasible alternatives. The feasibility of a solution by this method is guaranteed on account of earthwork costs. A method may generate more cut and fill quantities than if a feasible search-bound approach is not applied.

The paper did not show whether binary or real number representation methods were used. On the other hand, details on the design of both crossover and mutation operators were not given, which made understanding the algorithm’s behaviour too difficult. Furthermore, it was mentioned that both crossover and mutation were important to produce a solution without detailing whether the two operators’ designs were based on standard genetic algorithm or they were problem-oriented, and therefore comments cannot be given without knowing the nature of the two operators.

The study showed that a population of 600 individuals and more produced no difference in the efficiency of the algorithm and accordingly a population size of 700 was adopted. At each generation, 90% offspring were combined with the best 10% of the parent individuals for breeding in the next generation. The crossover and mutation probabilities were 0.55 and 0.03, respectively. The population was converged near to iteration 500 with substantial improvement within the first 150 iterations. No analytical information was provided (e.g. a graph or chart) to show the behaviour of the solution during the convergence (evolution) process.

The unfeasible solutions were treated using constant static penalty for every constraint violation. This approach could help to easily weed out the unfeasible solution. Meanwhile, it was difficult to distinguish between two solutions with different levels of violation. For instance, this approach applies the same penalty on two solutions, one with 4 degree gradient violation and the other with 2 degree.

It is important to mention that the study showed that the GA techniques provide a flexible analytical tool over other conventional optimisation techniques to solve vertical alignment problems with different parameters.

It is worth mentioning that an approach similar to the definition of vertical feasible bound is adopted in this study for point elevation generation of the initial population. Details are given in the experimental work presented in Chapter 5.

3.2.2 HORIZONTAL ALIGNMENT MODELS

3.2.2.1 Calculus of Variations (CoV) Based Model

CoV is a mathematical analysis tool that can be used for non-linear minimisation principles on infinite dimension function spaces (Olver, 2010). A simple example to define calculus of variations is to find the shortest length of a curve between two termini points. The answer to this problem, if no constraints are defined, is a straight line (Wikipedia). This is also considered similar to a typical highway alignment problem (Olver, 2010; Wan, 1995). The method requires a function with at least one differentiable ordinary sub-function (local costs for highway alignment problems). Each sub-function may represent a component or a set of variable components of the problem. The function is usually represented by the local cost components (sub-functions) of which the overall objective value is either maximised or minimised.

Shaw et al (1982) applied the optimum curvature principle (OCP), which was developed by Howard et al (1969) and derived from the calculus of variations, to horizontal alignment optimisation of an expressway in South Florida between two known end points. The optimisation was based on cost minimisation (right of way, earthwork, and construction costs). A cost surface was created and the resulting route traversed along low-cost fields. It was mentioned that South Florida is a relatively flat area and the earthwork cost was considered as sub-base soil rather than a vertical alignment. This consideration kept the problem as a two-dimensional rather than a three-dimensional problem.

The problem consisted of two parts: the engineering part, which dealt with the horizontal alignment selection, and the mathematical part, which dealt with the numerical solution of ordinary non-linear differential equations for the two-point boundary values (the two fixed locations) representing the mathematical statement of the OCP.

OCP equalises curvature at each point of an optimum route to the percentage rate of change. A set of routes with various starting angles is generated. OCP finds the sub-optimum for each starting angle, and from the set of sub-optimum solutions the best is selected. A solution that terminates at the end point was greatly affected by the starting angle of the initial solution, which might have exhausted the optimisation process. The study showed the successfulness of the method to guarantee a smooth solution.

3.2.2.2 Neighbourhood Search Heuristic Method

A two-stage design approach was developed by Lee et al (2009) to solve a horizontal alignment of a highway. A neighbourhood search-based optimisation-oriented heuristic approach was used at the first stage to select a “good” piecewise linear horizontal alignment, whereas the second stage was dedicated to the alignment adjustments and satisfying code requirements. The problem was formulated so that the solution has a minimum possible length, passes by the weighted control areas, and avoids the restricted ones. The lengths of the piecewise linear tangents were constrained in the model for proper horizontal curve fittings. The model then used the objective function value to compare the new solution with the previously generated one. The approach generates only one candidate at the start and it is being improved over successive iterations.

At the first stage, the heuristic yielded an approximate horizontal alignment represented as a piecewise linear line. The initial solution connected the two termini points through intermediate intersection points (IP). The linear line segments were improved iteratively by imposing large changes. The method, in addition to moving the current IPs to randomly selected positions, randomly added and/or deleted IP points to alter the number of line segments. This stage attempted to find a quick and good approximate alignment, whereas the second stage was more about fine tuning the alignment by making small adjustments to the piecewise linear line.

The objective function was formulated based on a weighted sum of the piecewise linear line segment lengths and penalty terms for both control and restricted areas. A soft penalty approach was applied according to the depth of penetration into the restricted areas and according to the distances between the alignment and the weighted control areas. This approach enhances the heuristic to distinguish the quality between different solutions.

Although the adopted approach produced a smooth and realistic alignment, the solution was not more than a local optimum due to the following reasons.

1. The model generated one candidate as an initial solution, which might not have been sufficient to search enough areas for alternative solutions.
2. The total length of the piecewise linear line was considered as the alignment length, while the actual alignment length including the curves was neglected.
3. The only critical cost component that was included in the objective function was the length cost, while other important costs such as location were ignored.
4. The constraint areas were represented in rectangular shapes.
5. The model was checked only over a simple control and restricted areas as constraints. No difficult study areas with different land categories were presented to demonstrate the efficiency of the model.

3.2.2.3 A Non-linear Mixed Integer Model by Easa and Mehmood (2008)

Mixed Integer Non-linear Programming (MINLP) refers to mathematical programming with continuous and discrete variables and non-linearities in the objective function and constraints (Bussieck and Pruessner, 2003).

A model for highway horizontal alignment design was developed by Easa and Mehmood (2008), based on an approach for safety. The new method was designed not only to implicitly consider the safety through code requirements, as with the traditional methods, but to include actual collision experience as well. The model was formulated as a non-linear mixed integer problem and solved by a pre-developed software package (LINGO) to obtain a global optimal solution.

The model decision variables were defined as PI locations, radii of curves, design superlevations, curve lengths, tangent lengths, and lane and shoulder widths. The model was formulated to find the decision values that maximise safety by minimising the mean collision frequency subject to geometric, collision (functional), and physical constraints.

Easa and Mehmood mentioned that collisions on horizontal alignment curves could be recognised as a significant safety problem. They referred to experience collision rates on curves as being three times higher than those on straight sections.

In the study, annual collision frequency on horizontal curves was estimated based on curve radius and tangent length. It was also shown that tangent length above a certain length may increase collision on the curve ahead. Therefore, the collision modification factor (CMF) was suggested to adjust the collision frequency to reflect the geometric characteristics of the curve more accurately.

The model is good to optimise the design elements of horizontal alignment. However, the capability of the model is quite limited for real alignment optimisation. First, the number of IPs were specific as input, while a free search is required to configure a realistic alignment regardless of the number of IPs. On the other hand, the objective function was associated with optimum element dimensions that minimise collision frequency, while highway design elements are not only functions of accidents but are also functions of other cost components as well (e.g. construction costs). Furthermore, the location of the alignment was not related with the process and the study did not demonstrate producing alignments with backward bends. No explanation was given on how LINGO software works. Finally, the solution approach was applied on horizontal alignment while vertical design elements also have impacts on collision frequency.

Therefore, the model would be useful for collision evaluation on pre-optimised highway alignments more than using it as a safety-based model for highway alignment optimisation.

3.2.3 THREE-DIMENSIONAL HIGHWAY ALIGNMENT MODELS

3.2.3.1 Calculus of Variation (CoV) Technique

Chew et al (1989) presented a CoV-based model for the three-dimensional highway alignment problem. The highway alignment was configured through the employment of a three-dimensional cubic spline. Chew et al’s work was an extension to a previous model by Goh et al (1988), which was developed to optimise vertical alignment through the DP method. Chew et al stated that DP uses coarse grids and produces a rough and approximate suboptimal solution due to its requirement for a large memory capacity. Therefore, they did not extend the work with the same technique. They proposed that DP could be useful to locate a band of interests (corridor) within a feasible area and then to refine the solution using another discrete technique.

Chew et al, in their 3D model, integrated earthwork and pavement costs as the objective function of the problem. Vertical grade, horizontal and vertical curvature, and inaccessible regions were considered constraints of the problem.

The characteristics of Chew et al's model and the solution quality that the model produces are summarised below.

1. The way that the problem was handled was complex.
2. The solution alignment was smoothed without giving consideration to the design requirements for radius, degree of the curve, design speed, and sight distances. Therefore, the resulting solution cannot be considered as a realistic alignment.
3. Required huge computational time.
4. The technique cannot hold non-differentiable cost components.
5. The horizontal and vertical alignments were considered in two independent stages and therefore the solution was suboptimal.

3.2.3.2 De Smith, 2006 Model (Network Optimisation)

This model was formulated to find as short as possible optimal path (length cost minimisation) for road, rail, and pipeline routes under the gradient and curvature constraint considerations. The method suggested using Spline for handling the curvature for both horizontal and vertical alignments. The format of the study area was made using square grid cells where elevation data were provided. Direct neighbour search was used to find the best square grid neighbour. This approach required, for each square grid finding, a search in eight existing directions. The method produced a rough path, while smoothing was obtained using the Spline method.

The adapted approach, as a technique to find the shortest path with curvature and gradient consideration, was not enough to produce a good solution. Moreover, the critical cost components (e.g. location, user, and environmental costs) and constraints (e.g. horizontal and vertical curve length for sight distance requirements) were omitted from the model formulation. The search procedure was restricted to neighbouring square cells and did not guarantee the exploration of the entire search space. Therefore, the obtained solution could not be considered more than a local optimum solution. The approach could be more suitable for pipe line routes due to the fact that pipe lines do not require the same strict requirements for curvature and gradients as highway alignments do.

3.2.3.3 Cheng and Lee, 2006 Model (Neighbourhood Search Heuristic)

A mathematical model for 3D highway alignment was proposed by Cheng and Lee (2006) to handle grade heavy vehicle speed, points of interest (control points), and restricted areas (no-go areas). The model employed the neighbourhood search heuristic technique and tried to take into account the conventional highway alignment design elements (e.g. tangents, circular and parabolic curves) including clothoid (spiral) curves for horizontal transitions. The model was formulated into two subsequent stages: horizontal alignment optimisation and then the vertical profile in another independent stage.

The model adopted the random search method for the IP locations and line segment lengths. The search consisted of either adding a new IP or deleting one or by moving an existing IP point. Later on and in a different process the design elements (e.g. tangents and curves) were inserted. Length cost and the penalty items were regarded as the total cost of the horizontal alignment. A weighted penalty approach according to the significance of the control and restricted areas was considered.

In the second stage, the vertical alignment was searched based on the vertical intersection point (VIP) elevations. The neighbourhood search heuristic was used as a solution technique and the grade between any two neighbouring VIP points was checked for the speed limit and used as a constraint to the problem. Any solution with grade violation was then penalised. The resulting solution at each iteration was evaluated and a solution with less cost superseded the current one.

The resulting alignment was smooth and realistic according to the design element consideration. However, the resulting solution cannot be seen as a global one in two aspects. First, the technique used one single solution and this might not be enough for an efficient search. Second, the alignment generations in two independent stages are likely to produce a local optimum rather than a global optimum solution. No attention was given to curvature value and sight distance on both the horizontal and vertical alignments. Finally, the model was formulated based on length cost only, while excluding other critical costs like environmental and earthwork costs was not justified.

3.2.3.4 Genetic Algorithm-Based Models

Before the GA-Based models are presented it would be important to understand what is genetic algorithm and how it works so that the models that use GA techniques are better understood. Section A below gives a brief review about the GAs and Section B and C present the models that use the technique for 3D highway alignment optimisation.

A) A Brief Review About GA

Michalewicz (1999) stated that the traditional optimisation methods are exhausting and are relatively efficient to search in small spaces, but for larger and more complex spaces more intelligent techniques are needed. Stochastic algorithms like genetic algorithms (GAs) are among such techniques.

A GA is an evolutionary adoptive search technique inspired by the theory of natural selection and survival of the fittest (Beasley et al, 1993; Coley, 1999). It is a mechanism that mimics the genetic evolution of species (Andre et al, 2001). Holland invented the technique in the early 1970s aiming to import mechanisms of natural evolution into computer systems. This technique has been modified in order to suit different problems and applications in different environments (Davis, 1991; Mitchell, 1996). A variety of studies has proven that GA is an efficient tool for planning and optimisation problems.

Mathews et al (1999) applied GA to land use planning, Mawdesley et al (2002) used GA for construction site layout in project planning, Ford (2007) used GA for housing location planning, Jong (1998), Fwa et al (2002), and Tat and Tao (2003) used GA for alignment optimisation problems.

GA is the process of combining the elements of two solutions usually known as parents, then to mutate a single solution to produce the third solution known as the child, and then evaluating it. The fitter individuals, after the parents and offspring populations have been combined, have more chances to survive and breed in the next generation, whereas the poor chromosome should die off. The main stages of a GA process include encoding the solutions, initial population generation, genetic operators (selection, crossover, and mutation), a customised evaluation function to evaluate the fitness of the candidate solutions, and terminations (Davis, 1991; Mawdesley et al, 2002; Michalewicz, 1999).

In GAs, the search space of the problem is the environment, and a set of different alternative solutions to the problem is the population. Each solution in the initial population is encoded into a string called a chromosome. To find an optimum solution by a typical genetic algorithm, the chromosomes of a population should undergo the application of three different operators. These operators are known as selection, crossover, and mutation (Coley, 1999; Davis, 1991). Over many generations and under the effect of these operators, the good characteristics will mix and spread out within the genes of the individuals leading to the exploration of the more promising areas of the search space. After successive generations, depending on the efficiency of the GA (a well-designed GA), the population should converge to an optimal solution (Beasley et al, 1993; Jong and Schonfeld, 2003).

The evolution of a population is greatly affected by the applied operators. Simple operators do not work well in some problems with a complex structure. The genetic algorithms' operators (crossover and mutation) are mostly problem specific and often they must be devised to fit in with the problem to efficiently facilitate the search process (Jong and Schonfeld, 2003).

Matthews et al (1999) and Beasley et al (1993) mentioned that GAs are not always guaranteed to find the global optimum solution to a problem, but they are efficient to search for “good enough” solutions. Therefore, for any particular optimisation problem, as stated by Tat and Tao (2003), a customised GA needs to be formulated.

The population of a well-designed GA will evolve over successive generations until the difference between the average fitness of individuals and the fitness of the best individual is minimal, by which the population is known to have converged. Both the gene and the population could be examined for their convergence. Beasley et al (1993) stated that “A gene is said to have converged when 95% of the population share the same value and the population is said to have converged when all of the genes have converged”. Accordingly, the average fitness will approach that of the best individual.

Different criteria have been suggested to terminate a GA process. For instance, Mawdesley et al (2002) reported that the operational performance of any genetic algorithm will continue until it confronts a predefined condition that terminates the continuity of the algorithm if the condition is to be satisfied (Mawdesley et al, 2002). Meanwhile, Jha and Schonfeld (2000) suggested that the search can be stopped at any time when the improvement in the objective function becomes negligible (Jha and Schonfeld, 2000). Other criteria could also be possible.

GA has been proved efficient for continuous search spaces, and when the objective function is non-differentiable (as is the case with the highway alignment problem). More importantly, the evolution with GA is based on a pool of candidates rather than a single solution (as is the case with some optimisation methods, as mentioned in the review of the models in the existing literatures). This is expected to find better solutions and reduce the chance of getting trapped at local optima.

B) The GA-Based Model

The very first GA-based model for highway alignment optimisation in 3D space was developed by Jong (1998). The search employed location cost, length costs, user-related costs, and earthwork costs in the objective function for alternative evaluations.

Jong represented the alignment by a series of intersection points (IP) in a three-dimensional space. These intersection points were distributed on vertical cutting planes located perpendicularly at evenly distributed distances along a straight line connecting the start and end points of the alignment (Figure 3.1). The intersection points were defined by X, Y, and Z coordinates. A piecewise linear trajectory alignment in 3D space was then generated by connecting the start and end point of the alignment through the intermediate IPs. Next, the two alignments were generated in two different procedures. First, a rough horizontal alignment (connected tangents) was obtained from the orthogonal projection of the 3D piecewise linear trajectory on the horizontal plane. Horizontal circular curves were inserted at each horizontal intersection point location (HIP) to form a smooth transition between the successive tangents. Penalties were used over the violation in curve length and/or tangent distance limitations. Second, a rough vertical alignment was also generated from the alignment projection on a vertical plane surface and then smoothed by inserting parabolic curves at each intersection point. The vertical plane was represented so as to cut the horizontal alignment along the centreline of its length. The chromosome representation, the GA operators that Jong used, and the model characteristics are briefed below.

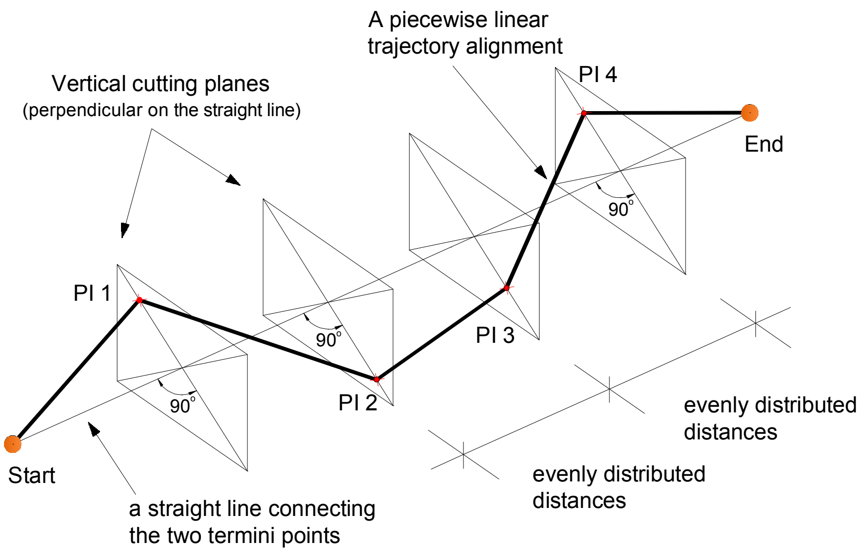


Figure 3.1: A piecewise linear trajectory 3D alignment representation by Jong (1998)

a. Chromosome representation

Each solution candidate was represented through encoding the X, Y, and Z coordinates of the IPs, including the start and end points, into the chromosome genes. The genes were considered as decision variables on which the objective function formulation was based.

The initial population was then generated using five different approaches, as follows.

1. Intersection points lie on the straight line connecting the start and end points: this was expected to reduce the length-dependent cost.
2. Intersection points lie randomly on the perpendicular planes with random elevations: this was to produce different intersection point locations.
3. Intersection points lie randomly on the perpendicular planes with elevations as close as possible to the existing ground elevations: this was expected to reduce the amount of earthwork.
4. Intersection points scatter randomly within the study region with random elevations: this was to carry useful information to produce backtracking alignment.
5. Intersection points scatter randomly within the study region with elevations as close as possible to the existing ground elevations: this was to produce backtracking alignment with minimum possible earthwork.

b. Selection

The selection scheme of Jong's model was modified over the selection algorithm proposed by Michalewicz (1996). It selects n_r parents to breed and the resulting offspring were replaced by n_r chromosomes in the parent population. The selection probability of an individual was defined as a non-linear function of a user-defined parameter, q. The scheme gives more chance to a better individual for being selected. The function was as below:

$$P_k = \frac{q\,(1-q)^{k-1}}{1-\,(1-q)^{n_p}}$$

Where P_k is the selection probability for the k^{th} chromosome in a ranked population, q is a user-defined parameter [$q \in (0,1)$], and n_p is the population size.

c. Crossover and Mutation operators

Four problem-specific crossovers and four problem-specific mutations were devised. Jong stated that simple GA operators did not work well and this was attributed to non-binary representation of the problem and indirect relation of the genes to the 3D alignment. The relation was considered indirect because the horizontal and vertical alignments were generated by independent curve fitting procedures and were not interpreted by the genes. Simple crossover, two-point crossover, arithmetic crossover, and heuristic crossover were developed as crossover operators while uniform, straight, non-uniform, and whole non-uniform mutations were used as mutation operators.

d. Characteristics of the model

The model was successful in producing a realistic alignment. However, this does not mean that the model is perfect in every aspect. Despite the GA being well looked at to solve the problem, the following comments are believed to be the defects of the solution algorithm.

1. Generating five sets of population might deteriorate the search in some specific cases. For instance, generating straight alignments between the termini points might never be satisfied especially at hilly or difficult terrain areas and there might not be a solution that would locate the straight line between the two termini points. Thus, such a solution will definitely die off through the successive generations instead of working as a good knowledge provider for the new offspring. This is also true for generating alignments as close as possible to the natural ground elevation. The latter could produce very steep grades, especially at cliff locations or deep valleys and might also restrict the distances required for

the curve fitting and sight distance. Such an approach could overburden the search.

2. The search is almost restricted by defining vertical cutting planes located at evenly distributed spaces on the straight line connecting the two end points. Although there is a specific operator that generates random locations for the intersection points (the decision variables) within the whole study area, most of the operators perform the search within the limits defined by the boundary of these vertical cutting planes. Restricting the search space may lead to missing out on exploring some good locations.
3. The introduction of the vertical cutting plane approach led to the design of a special algorithm to find the intersection point coordinates within the planes. Two complex procedures were specifically formulated to determine the origin of the planes and then to calculate the IP locations within the plane.
4. The study did not explain whether the eight crossover and mutation operators perform the task together or separately. However, this was another sign of complexity. No individual test on each operator was shown and therefore, it is difficult to recognise which was better or which one was dominating the search.
5. Transforming the generated piecewise linear trajectory line into a smooth alignment was done through two independent curve insertion procedures: horizontal curve fittings and vertical curve fittings. This procedure was restricted by the tangent distances and therefore special algorithms were developed to handle the case and to fit suitable curve lengths at the intersection point locations.
6. The model did not introduce any new or novel ideas related to the design approach of highway alignment to reflect the modern technology development. The model formulation was totally based on the traditional (existing) geometric design approach of highway alignment. Most of the traditional steps to design a highway alignment were embedded into a GA-based search model to find a good highway alignment solution. The aim of the research was just to speed up the

design process and to find better solutions through evaluating more alternatives than the manual methods do.

7. The model optimises highway problems based on its own factors. The effect of access demands from the land uses of the area under consideration was not included.

The same model was extended by Kang et al (2007) and Kang (2008) to optimise a three-dimensional highway alignment within a simple road network including the determination of the start and end points. The model was similar to Jong's model in dealing with the main highway alignment, other than in the following aspects.

1. The model dealt with a three-dimensional highway alignment within a simple road network.
2. Transition curves for the horizontal alignment were also included.
3. The highway alignment termini points were unknown and had to be determined.
4. The aim of the study, other than a three-dimensional highway alignment optimisation, was also to improve the traffic performance of the road network with the addition of a new highway. For this purpose, the traffic impact on different alternatives was evaluated.
5. Using different approaches to define a specific area of interest (confining the search boundary) for the search process, as well as adopting different techniques to handle the constraints. Some explanations of these approaches are presented below.

Kang in his model made the search rather restricted and critical to the problem. A user area of interest was defined within which the optimum solution was found. The horizontal search in the XY plane was restrained by the distances resulting from horizontal intersection of the vertical cutting planes, as with Jong's model, with the boundary of the user's area of interest (Figure 3.2). The vertical search was also restricted by the distances resulting from the intersection of maximum allowable tangent grades with the vertical cutting planes. This approach, which was called feasible gates (FG) and was first developed and described by Fwa et al (2002), was believed to enhance the computation efficiency and solution quality of Jong's model.

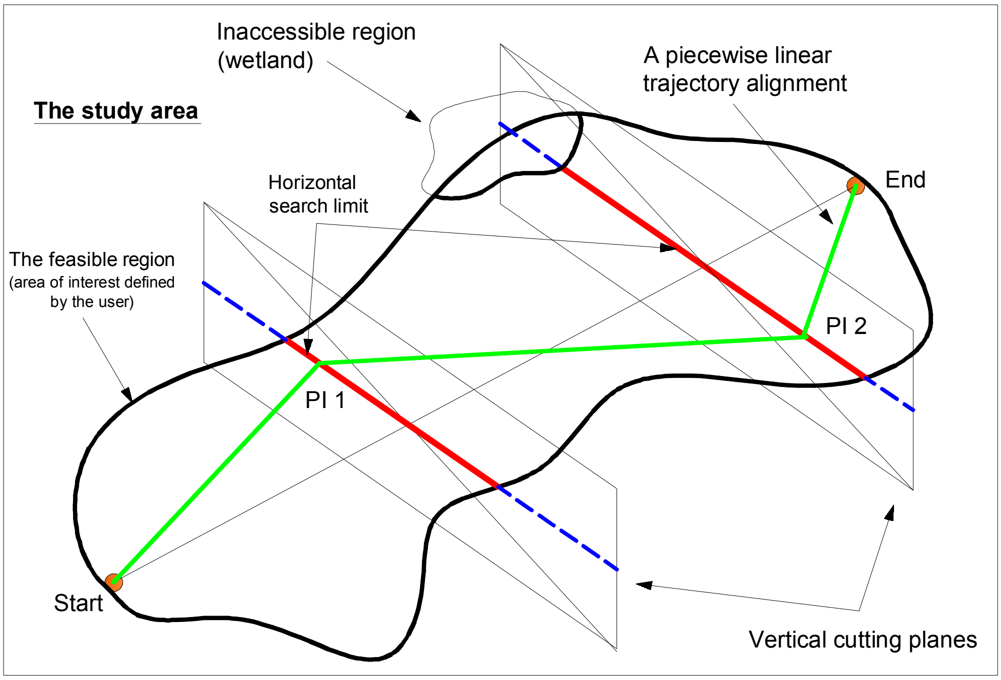


Figure 3.2: Feasible gate approach for horizontal highway alignment by Kang (2007)

Giving preference to an area for search by a user without having evaluated the entire search space may lead to obtaining a solution that is only optimum for the specified area. Such a definition by itself is a complex task and it might not be easy to recognise a feasible corridor within a complex configured area. Therefore, user preferences for defining an interested area are unlikely help to produce a global optimum solution. This approach leaves no preferences to the areas which are located outside the defined area. Ignoring these spaces means that various alternatives are ignored too.

Another GA-based model, which seems to posses the same approach and principles as Jong's model, was developed by Tat and Tao (2003). No clear information was given in the study and therefore it was difficult to give comments on the model's different aspects.

C) GA-Based Models with Geographic Information System (GIS)

Following Jong's model, as presented in the sections above, a group of researchers continued to improve the efficiency of the model. In this regard, the possibility of integrating the model with a GIS was investigated.

GIS was used to precisely calculate the areas of irregular land spaces, consider environmental impacts such as wetlands and flood plains, and hold some cost-related elements. Jha and Schonfeld (2000) mentioned that some highway alignment costs are geographic-related costs and therefore a GIS was exploited to obtain these costs. The costs which were considered sensitive to geography of any highway alignment were right of way, earthwork, and environmental costs.

The first GIS integration was developed by Jong et al (2000) and extended by Jha and Schonfeld (2000) and Jha et al (2001). They suggested the use of GIS for the following advantages.

1. Can use real maps.
2. Precisely represents the environmentally sensitive areas which are affected by the alignment location.
3. As a geographic data warehouse, GIS can make geographic information handy for the model.
4. Some calculations like land parcel areas, which are affected by the right of way, can be calculated within the GIS for precise land acquisition cost calculations.

The calculation of the total location-dependent cost was automated through the use of GIS software (e.g. ArcView GIS) (Jong et al, 2000) instead of manual programming within the model formulation as was the case in Jong's model (1998).

The GIS creates a buffer around the centreline of the alignment and superimposes the boundary of the alignment on a real map with different land parcels. The land parcels where the alignment passes through can be identified by GIS and the associated areas and costs are calculated. The results are then sent to the GA model to complete the search process. The calculation process requires that the geographic data be available in a digital format, including digital maps.

It can be noted that the GIS does not help the algorithm to perform a more efficient search other than providing spatial information for more realistic and precise representation. A GIS, as stated by Jha and Schonfeld (2000), can provide precious input to a highway design optimisation model besides computing geographically sensitive costs.

During the search process, a continuous exchange of data would be required between the two models. Therefore, the integration of GIS with a highway optimisation model, due to the additional computation necessary in the GIS environment, will significantly slow down the search process. The difference in the environment of the two models adds some load to the data-processing operation and makes it harder than if they were the same. These differences require pre-processing the GIS data to make them compatible with the optimising model (Jha, 2002; Jha and Schonfeld, 2000).

More details on the methodologies, principal concepts, adopted GA with their basic and problem-specific operators, integrating the GA model with GIS, cost functions, and improving the efficiency of calculation of a genetic algorithm-based model for highway alignment optimisation could be found in the works carried out by Jong et al (1999), Jha et al (2001), Jong et al (2000), Jha (2003), Jong et al (2003), Tat and Tao (2003), Kim et al (2005), Jha et al (2006), Jha and Maji (2007), Kang et al (2007), and Kang et al (2009).

3.3 A BRIEF OVERVIEW

The existing models for highway alignment design were reviewed and it was found that they were all based on the traditional (existing) design approach for highway alignment design. No new design approach was found in the literature and the existing methods were found to share the same designing principles. In general, the aims of these studies could be summarised in three points:

1. To automate the process

2. To speed up the process
3. To evaluate more alternatives than the traditional methods can do in order to seek out better solutions.

On the other hand, as also agreed by Goktepe et al (2009) that these models differed from each other by the optimisation techniques, the defined constraints, and the cost components involved in the objective function formulation. Jha (2002) stated that the approaches in the literature, other than GA, are traditional approaches and were unable to solve realistic problems due to oversimplifying the assumptions necessary to employ the algorithms. Thus, according to Jha, the resulting solutions lacked reliability.

The aim of this research study is to investigate the possibility of improving the highway alignment development process through proposing the notion of station points as a new design approach for highway alignment development. Therefore, these models were found useful to identify the main components of a 'good' model. Accordingly, the possibility of combining and integrating some of the methods, ideas, and concepts of the existing models were investigated to produce a model that employs the notion of station points for highway alignment configuration. The sections below outline the main structure of the model formulation proposed for this research study.

3.4 THE PROBLEM FORMULATION OF THIS RESEARCH STUDY

As stated in the introduction section of this chapter, the revision of the existing models for highway alignments was to find the components of a good model rather than finding out how the used techniques work or to compare them to each other.

By the aim of this research, an approach is suggested to improve the development process of highway alignments. Therefore, taking advantages from the background principles of highway alignment development, as presented in Chapter two, and from the existing models, the following elements were found to be essential for a good model formulation. These components are expanded in the succeeding sections.

1. A search technique to perform the optimisation process. This is selected to be the genetic algorithms.
2. Format the study area in such a way that the spatial and geographic data can easily be represented. For this purpose a GIS data model format was reviewed to understand how well a study area can be formatted.
3. The formulation of the most critical cost components to the problem that could help to produce an answer to the defined aims and objectives. This is important to evaluate the candidate solutions so that the most adapted solution, according to the associated criteria, is determined.
4. Posing constraints so that the solution is reliable.

3.4.1 GA AS A SEARCH TECHNIQUE

Highway alignment is a constraint optimisation problem. It is characterised by sophisticated non-linear constraints and objective function (Goktepe et al, 2009). It has been proved that GA is efficient to solve complex problems; therefore, this research study also suggests employing the method as a tool to perform the search. More details about genetic algorithms were given in Section 3.2.3.4 (A) of this chapter.

3.4.2 THE STUDY REGION FORMAT AND GEOGRAPHIC DATA

Highway alignment optimisation requires enormous geographic data. For this purpose, in order to increase the operational performance of the proposed model, the GIS-related features are reviewed so that the study boundary is formatted on similar bases.

GIS is a computer system known for creating and managing spatial data and associated attributes, and is capable of integrating, storing, editing, analysing, and displaying geographically-referenced information (Cowen, 1988; Demers, 2005; Heywood et al, 2006; Khatib et al, 1999; Longley et al, 2001; Maguire et al, 2005; Rhind, 1988; SSNDS, 2006).

Transportation planning is one of the planning fields that has benefited from GIS technology. The application categories that GIS can deal with are road design and alignment optimisation, highway mapping, traffic analysis, and shortest route finding (Alterkawi, 2001).

3.4.2.1 Spatial or Geographic Data

Identifying physical locations of objects and their properties on or near the earth is known as geographic data (Galati, 2006).

Spatial data represent information about features' positions, relations with other features, and information about non-spatial characteristics at that position. Simplification of data is required before they can be stored in a computer. The most common method for this simplification is to break down the geographic features into the three basic entity types known as points, lines and areas to represent objects like buildings, rivers, and forests, respectively. All these real world features are stored in a GIS by means of two representation data models: raster or vector (Heywood et al, 2006). These data models are different in structure; however, they have both been used to represent real world features (Galati, 2006).

i) Vector data structures

The feature geometry in a GIS vector model is exploited to create objects that consist of points, lines, or polygons or a combination of these feature geometries (features with recognised boundaries like points and roads) (Galati, 2006). By this method, points are represented by the X and Y coordinates, lines as a set of the X and Y coordinates connected in order; and areas are represented by a set of X and Y points connected in a closed boundary (Demers, 2005).

ii) Raster data structures

The second major GIS data type is digital images represented by a grid of valued pixels or cells and they are known as raster data. Raster data is a model used to depict the earth's features through digital images. To depict continuous data like imagery where the feature boundaries and point information are not well recognised, raster representation would be efficient. Feature geometries of a vector data model compared with a raster data model are presented in Table 3.1 (Galati, 2006).

Table 3.1: Feature geometries of raster and vector data

Feature	Vector Model	Raster Model
Point		
Line		
Polyline		
Polygon		

The raster data model uses two-dimensional array grid cells (pixels) to represent real world objects. Each cell can hold an attribute or multiple attributes and both binary and floating point methods are common as encoding schemes for their representations (Longley, 2001). Geographical feature representation is found by the number of colours and image type in which they determine the properties and appearance of the grid pixels. Figure 3.3 below shows three different property types represented by raster data using three different colours (Galati, 2006).

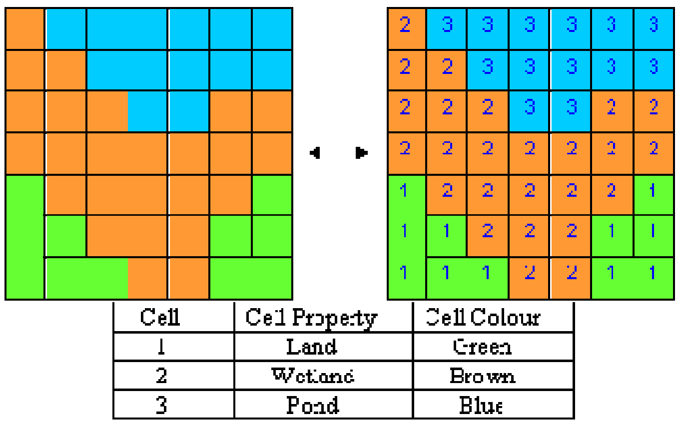


Figure 3.3: Representation of different property types using raster data type

3.4.2.2 Layers in GIS

One of the GIS features is using layers. Layers in GIS gather geographic information of the same type, which facilitates editing data and building relations among the entities (Longley, 2001).

Figure 3.4 shows how different spatial data in a GIS are separated in different representative layers (SSNDS, 2006).

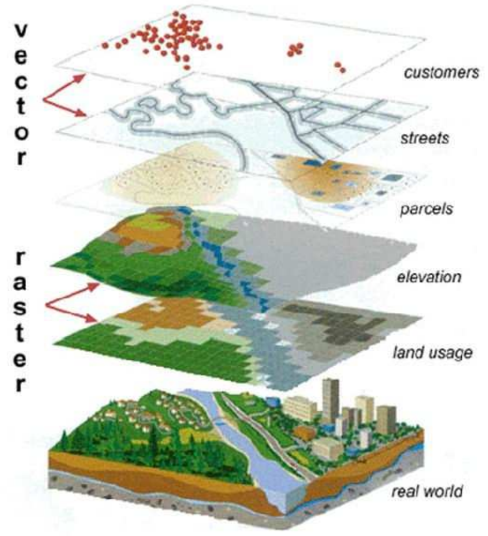


Figure 3.4: Geographic feature layers in a GIS (Source: http://ssnds.uwo.ca)

3.4.2.3 Grid System in GIS

To find, locate, and calculate distances and directions on the earth, a coordinate system would be essential. If the feature to be located is confined to the spherical earth or its reference globe, the geographic coordinate system that uses latitude and longitude would be useful. If two-dimensional maps are dealt with to locate objects correctly, a rectangular or plane coordinate system would be used (Demers, 2005).

The rectangular coordinate system consists of an abscissa (X axis) and an ordinate (Y axis). The abscissa originates at (0), called the origin, and extends to the right for positive values and to the left for negative values. The ordinate allows movement vertically up if the value is positive or down for negative values from the same origin. Using this system, both the X and Y values are needed to locate any point or object on the earth's ground surface (see Figure 3.5 for detail).

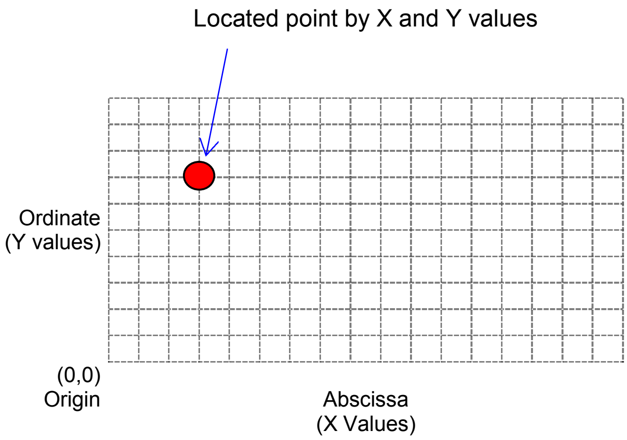


Figure 3.5: Cartesian coordinate system

3.4.3 THE COST FUNCTION

This research tests the practicality of using the station point design approach (SPDA) to improve the highway alignment development process and investigates the effect of junctions on the optimum alignment location and configuration. Therefore, the cost function formulation is based on the components that are thought critical to the desired goal rather than intending to produce a solution with concise cost value. These components are considered as length-dependent costs (construction costs), location-dependent costs (land costs), and earthwork costs. These costs are discrete and have different impacts on the solution, which may help to identify the characteristics of the solution by which the applicability of the proposed approach is evaluated. Using these cost components, the model performs the search based on length minimisation, avoiding high land costs, and minimises the costs of earthwork material. Each of these costs directs the search in a different way.

It is worth mentioning that the user costs are also formulated and used to test the sensitivity of the developed model against other cost components. The highway alignment development of this study is subjected to horizontal curvature, vertical curvature, and maximum gradient limitations in addition to imposing the alignment, as a user option, to balance the quantities of cut and fill materials.

On the other hand, when the model is used to deal with junctions, the link construction cost and the link traffic costs are used as the links' accessibility cost while considering a pre-defined minimum distance between the land use and the alignment as a proximity constraint.

3.5 SUMMARY

The existing models for highway alignment optimisation were reviewed. The following statements summarise the main differences between these models and the one that this research proposes.

1. Most of the models in the literature optimise either horizontal or vertical alignment while this model is designed for 3D highway alignment optimisation.
2. The existing models for 3D highway alignment optimisation, other than the GA-based one, dealt with the horizontal and vertical alignments in two discrete stages that make the search produce a local optimal solution rather than a global one. Meanwhile, this study considers both alignments simultaneously.
3. Most of the models in the literature use different search approaches than GA.
4. Most of the studies in the literature formulated their models based on the traditional geometric design approach for highway alignment. Meanwhile, this study proposes a novel design approach which, instead of using the design elements to configure the alignment, uses station points along the centreline of the alignment as a key element for the stages of planning, design, and implementation on the ground.
5. None of the studies in the literature explicitly investigated the effect of access provision on the optimality of the highway alignment. Considering access for the land uses may transfer the concept of the optimisation from a single highway to a road network system. This study suggests three access-related issues: (i) the possibility of integrating the access problem with the concept of the proposed

station point approach, (ii) optimising optimum junction locations along the main highway alignment, and (iii) investigating the effect of optimum junction locations on alignment location and configuration. The latter will provide an inevitable argument for a new trend towards simultaneous junction-highway alignment optimisation.

The following chapters detail the approach, techniques, formulations, experiments, and the test results that are attached to the aims and objectives of this research study for highway alignment development.

CHAPTER 4

AN OPTIMISATION MODEL FOR TWO-DIMENSIONAL (2D) HORIZONTAL HIGHWAY ALIGNMENTS

4.1 INTRODUCTION

Highway alignment is a three-dimensional problem in space. Building a three-dimensional model to solve the problem needs sophisticated procedures to handle the elements that are relevant to the problem. Therefore, a model needs to be built step by step to ensure that all required elements at each stage are tested and incorporated, and the design and operational requirements are satisfied. Furthermore, the model needs to be tested on worlds with a known solution for verification and then to go forward for more developments.

In this chapter the possibility of building a GA-based model for two-dimensional horizontal highway alignments is presented. A new design approach to generate highway alignment is introduced and its applicability to real life problems is investigated.

This chapter starts by defining the new design approach for horizontal highway alignment development. The required format for the region of interest and the way of handling the information is explained. The alignment representation, cost components, and planning and design limitations are presented. Furthermore, different GA operators are presented, tested, and tuned on different world features. The test results are analysed and presented in different tabular and graphical forms. The chapter ends with the presentation of the best parameter values by which optimum 2D highway alignments, with the use of the proposed design approach, are obtained.

4.2 THE PROPOSED DESIGN APPROACH FOR 2D ALIGNMENT

The proposed design approach depends on highway alignment station points along the alignment centreline. This approach is inspired by the fact that any generated alignment by whatever method will finally consist of a series of station points and it will be implemented on the ground depending on those station points.

From a practical point of view, station points are locations defining point positions along the centreline of an alignment and more often are spaced at equal distances. The distance between the successive stations may vary but the most common distance is 20 to 25 metres. This distance may be less at curves where higher precision is required and can be more on straight and easy distances. Traditionally, the determination of these points is the result of the final alignment geometric design when the detailed information and maps are prepared to make the design ready for implementation (Uren and Price, 2010) (Allan, 2007). The traditional design method of highway alignment, in the first stage, deals with finding the values for the elements of highway geometric design to form the alignment, and then in the second stage the design is converted to stations along the alignment centreline.

The new design approach of this study tries to combine the above mentioned two stages in one development stage. The method tries to directly exploit the requirement for the station points at the start of the design stage. It suggests optimising horizontal alignment of a highway through station points. Station points along the centreline of alignment are used to define the alignment configuration. The station points' positions are defined by their X and Y coordinates along the centreline of the alignment. These station points (let it be N number) are then interconnected sequentially in the order that they appear to form the alignment. In this study it is assumed that the start and end points are known. Figure 4.1 illustrates the concept of the proposed approach for a 2D highway alignment.

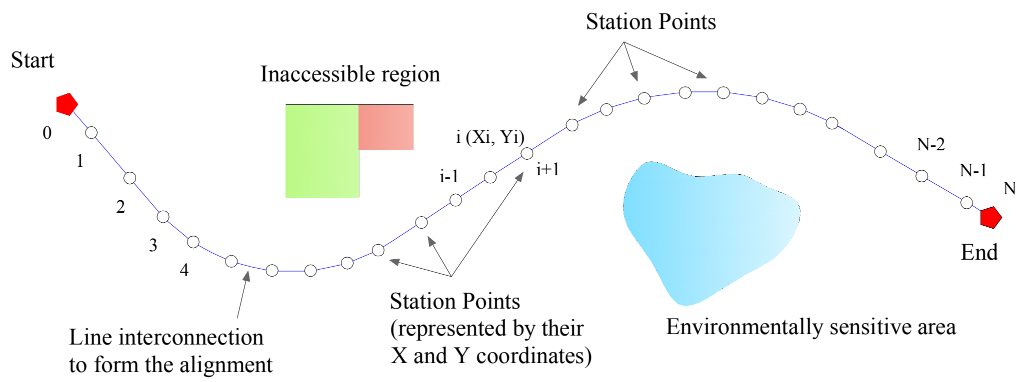


Figure 4.1: The new approach of station point for highway horizontal alignment configuration

An optimisation model is required to search for point locations that depict the best horizontal highway alignment. The final location of the station points is supposed to represent a solution with least fitness value according to the cost components assigned to the problem and satisfy the standard requirements and constraints imposed on the problem. it should be mentioned that the approach treats the bend segments the same way as the straight segments are treated and should provide the required bends without the need for any curve fitting process as the traditional design method does.

This study has suggested developing a mathematical optimisation model to handle the problem. The structure of the model is outlined in the sketch below where the main components of the model are shown.

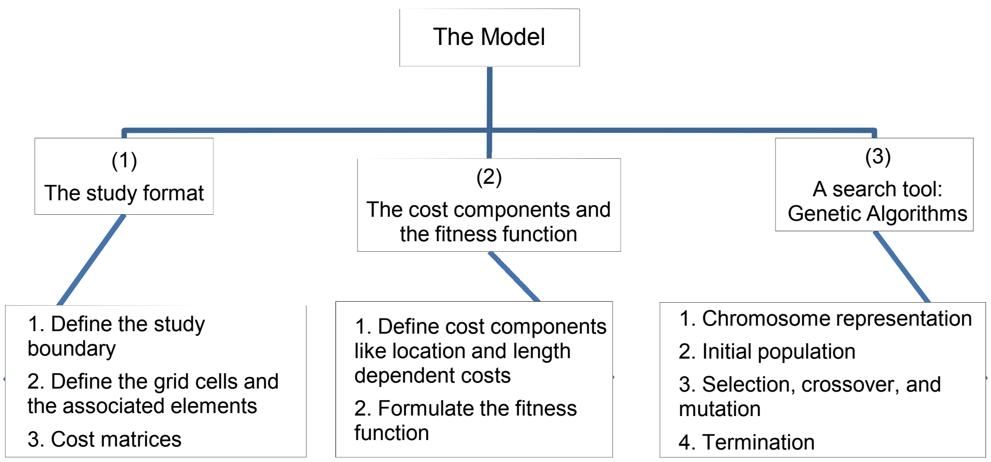


Figure 4.2: A sketch for the components of a highway alignment model

The following sections describe the step by step procedures to build a model for the horizontal highway alignment problem.

4.3 THE MODEL FORMULATION

4.3.1 THE STUDY BOUNDARY

The different land categories and elements of the study area, where building a new highway alignment is proposed, need to be defined and formatted in such a way that suits the model structure. The man-made features, environmentally sensitive areas, soil conditions, land use type, and land acquisitions are some example parameters that are associated with the data format of the study boundary.

In this research, the study area was considered as a rectangle defined by its width and height and divided into rectangular grid cells usually produced from a GIS model of the area under consideration. The size of the grid cells falls within the user preferences and depends on the desired accuracy. The smaller the grid sizes, the more precise would be the representation. Each grid cell may store several values which together define the cell and the way it behaves within the model. These values could be, for example, average land unit costs for 2D alignment problems and/or average land elevations for 3D alignments. For instance, an average land cost for a single cell is obtained from a set of land costs taken within the entire area bounded by the cell.

In this study and for the purpose of handling the land unit costs for a 2D highway alignment problem, a two-dimensional cost matrix was designed. The dimensions of the matrix were made compatible with the grid dimensions of the study area and were calculated as follows:

- *Let LUC be a two-dimensional Land Unit Cost matrix of dimensions n (number of rows) and k (number of columns), see Figure 4.3.*

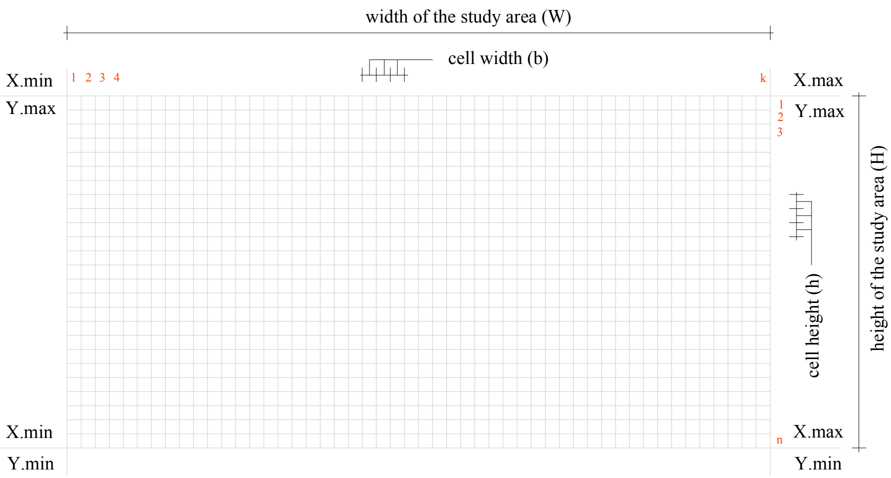


Figure 4.3: The grid cell format of a study area

$$n = \frac{H}{h} \quad , \quad k = \frac{W}{b} \tag{4.1}$$

➢ *Then the dimensions of the LUC matrix are (n,k)*

$$\mathrm{LUC}\,\{n,k\} = \begin{bmatrix} 1 & 2 & 3 & \dots & \dots & k \\ 2 & \dots & \dots & \dots & \dots & \dots \\ 3 & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & & & & \\ \dots & \dots & & & & \\ n & \dots & \dots & \dots & \dots & n,k \end{bmatrix} \tag{4.2}$$

The elements of the two-dimensional matrix represented the LUC values of the corresponding grid cells. The values of LUC were used to provide data for location dependent cost calculations of the objective function (details are provided in the objective function cost components). Figure 4.4 and Table 4.1 present an example of a two-dimensional study format with the corresponding unit costs required for a horizontal highway alignment model.

The study example is made for illustration only. It consists of nine different land parcels each with one or more than one characteristic that incur different cost values. The different land features of the example represent samples of real world cases and they are the most frequently-existing features in real worlds.

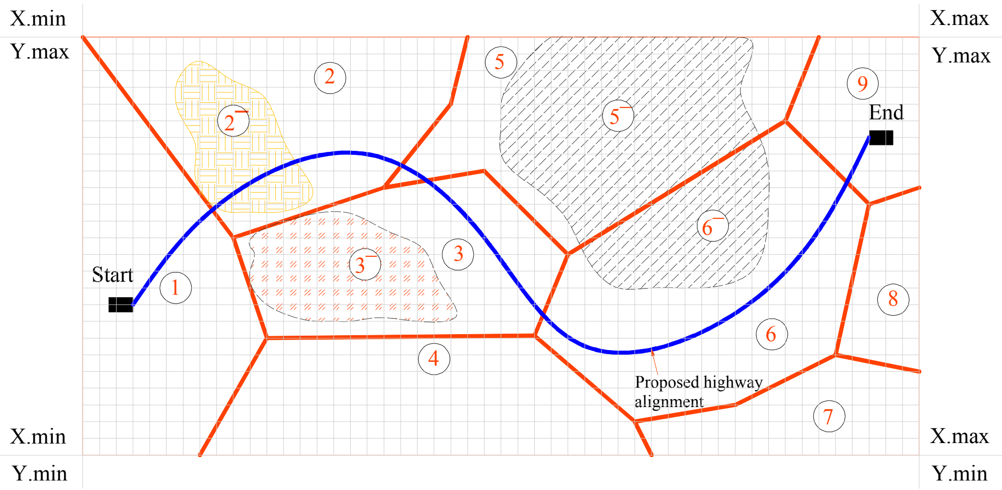


Figure 4.4: A study area with different land properties

Table 4.1: Definition of different land properties for a study area

Area Code	Description	Cost value
1	Government land parcels – Normal soil condition	0 or 1 or Relatively low unit location cost
2	Owned land – Normal soil condition	Cost of land acquisition
2^{-}	Owned land – with special soil treatment requirements	Cost of land acquisition + cost of special treatment for the soil
3	Government land – Normal soil condition	0 or 1 or Relatively low unit location cost
3^{-}	Government land – Political area	A very high cost: a user defined cost that makes the area inaccessible
4	Built up area – City	A very high cost: a user defined cost that makes the area inaccessible
5	Government land – Normal soil condition	0 or 1 or Relatively low unit location cost
5^{-}	Government land – Environmentally-sensitive features; e.g. forest or wetland	Cost of environmental impact (a user defined value)
6	Owned land	Cost of land acquisition
6^{-}	Owned land – Environmentally-sensitive features; e.g. forest, wetland, or lake	Cost of land acquisition + Cost of environmental impact (a user defined value)
7	Owned land – Normal soil condition	Cost of land acquisition
8	Owned land – Normal soil condition	Cost of land acquisition
9	Government land – Normal soil condition	0 or 1 or Relatively low unit location cost

The origin {O(X_0, Y_0)} and the upper right corner ($X._{max}$, $Y._{max}$) of the study area are either determined or predefined and any spatial object within the study area is then located by their X and Y coordinates. Based on this principle the coordinates of the highway station points were located using the Cartesian coordinate system. Figure 4.5 illustrates locating and representing an object and/or a station point within the study boundary.

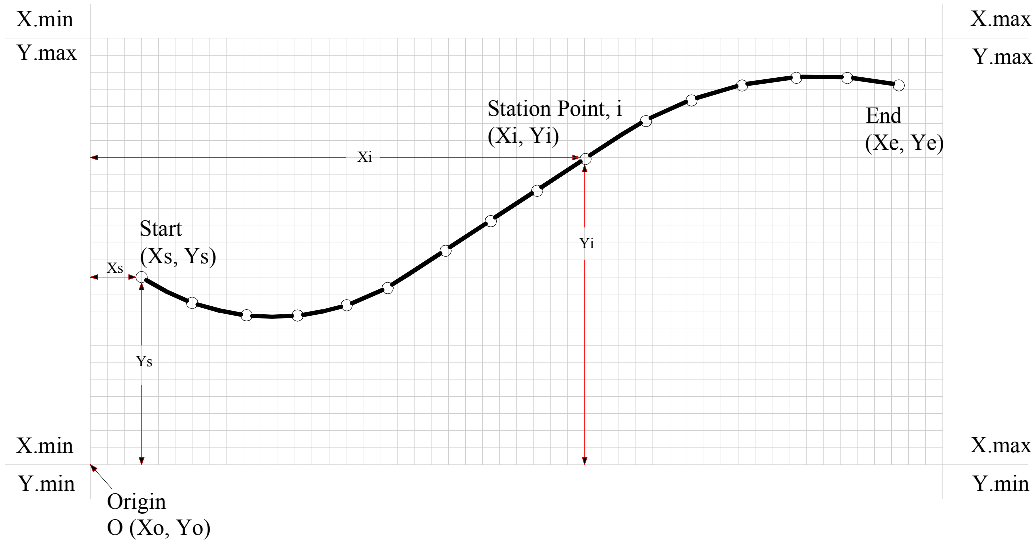


Figure 4.5: Location determination using the Cartesian coordinate system

4.3.2 THE COST COMPONENTS FOR A 2D HIGHWAY ALIGNMENT

In this study, the goodness of any alignment was evaluated in terms of cost. The lower the cost, the better is the solution. In general, many costs could be included and the optimum alignment should trade off among them. In this chapter, as the aim was to develop an optimisation model for a 2D highway alignment, the following cost components were therefore formulated and combined in a fitness function for alternative solution evaluation. These costs were related to:

1. Client or General Costs
    - Location costs (land acquisition, special soil condition, environmentally-sensitive area costs)
    - Length dependent cost (construction costs)
2. User Costs
    - Fuel consumption costs

- Travel time costs

User cost is listed here but not included within the fitness function. Its effect on the final solution is presented in Chapter 6, Section 6.10.1 and used within the model sensitivity tests.

The following sections give some details for the incorporated cost components were used in the 2D model formulation.

4.3.2.1 Length Dependent Costs (C_{Length})

In this study, the highway alignment construction cost (pavement cost: the cost of constructing the layers of a road pavement) was considered as length dependent cost. Equation 4.3 was formulated for this purpose:

$$C_{Length} = L \; x \; W \; x \; Unit \; Construction \; Cost \tag{4.3}$$

Where L is the total length of the alignment and W is the width of the pavement. The alignment length is considered as a function of the X and Y coordinates of the station points (decision variables) and it was calculated as:

$$L = \sum_{i=0}^{N-1} \sqrt{(X_{i+1} - X_i)^2 + (Y_{i+1} - Y_i)^2} \qquad \text{For all i = 0, 1, 2 … (N-1)} \tag{4.4}$$

Where N is the total number of station points used to configure the highway alignment in 2D space.

4.3.2.2 Location Dependent Costs ($C_{Location}$)

This cost may represent one or several location cost categories. Some features that are related to this cost in this study are land acquisitions, special requirements for poor soil treatments, and environmental features like wetlands, floodplains, and marshes. The location might also be man-made features like houses or buildings (or a city) or could be political or historical locations. Some of these locations are accessible and some are not. If any single cell within the study region boundary is inaccessible due to either political or historical or environmental factors, a relatively high unit cost is assigned to the cell representing that area. An alignment passing such areas possesses a very high fitness value and most probably dies off before it is being selected for breeding in the next generation. Moreover, when a location requires more than one consideration (e.g. land acquisition and soil treatment), the land unit cost would be the total of both and is attached to the associated cell. Also, when a cell represents an environmental area like floodplain or marsh, a user defined cost as to whether to preserve this area or not is associated with the cell.

The sum of the categorised unit costs was regarded as LUC and was represented as below:

$$C_{Location} = \sum_{k=1}^{p} l_k \; x \; W \; x \; LUC_k \qquad (4.5.1)$$

Where: $C_{Location}$ is the total alignment location cost; l_k is the length of the alignment located in a grid cell (k) with a specific cost value; W is the width of the area covered by the road; LUC_k is the Location Unit Cost of the cell (k); and p is the total number of cells that the alignment passes through (see Figure 4.6). Thus:

$$L = \sum_{k=1}^{p} l_k \tag{4.5.2}$$

Where L is the total length of the alignment.

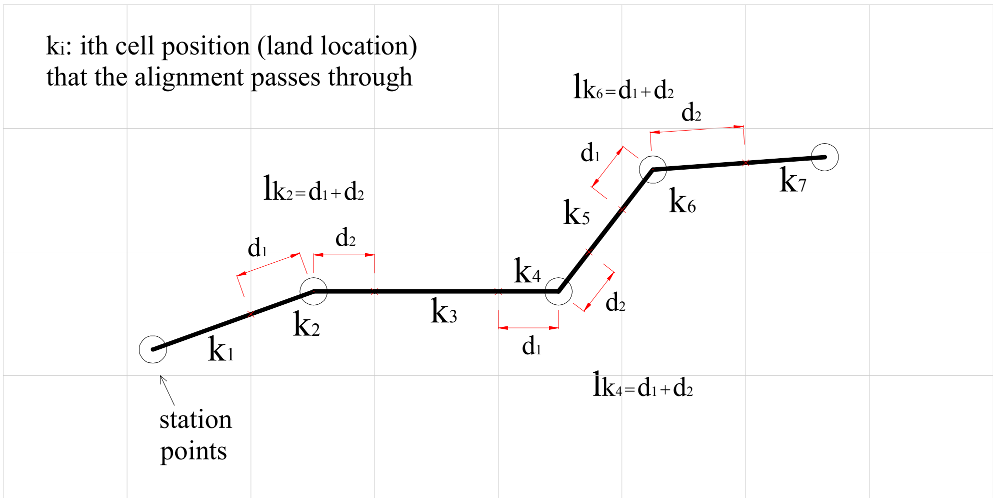


Figure 4.6: An illustration for alignment length related cell locations

LUC_k, as mentioned above, may represent one or more than one unit location cost:

$$LUC_k = \mathrm{LAUC}_k + \mathrm{STUC}_k \tag{4.5.3}$$

Where $LAUC_k$ is Land Acquisition Unit Cost of cell k (unit cost/m^2) and $STUC_k$ is Soil Treatment Unit Cost of cell k (unit cost /m^2).

4.3.2.3 User Costs (C_{User})

User costs are the costs that the uses of the road alignment incur on the users including the drivers and the travellers. User costs take the form of three main items recognised as vehicle operation costs, travel time costs, and accident costs. The relations of user costs with the alignment design elements and the decision variables that represent the alignment solution are not as direct and explicit as location and construction cost components. This cost component is affected by many factors and therefore it cannot be precisely represented as it is thought to be (see Chapter 2, Section 2.10.1.4 for more details). The user cost of this study is only used to test the sensitivity of the model to other different costs and this is shown in Chapter 6, Section 6.10.1, Scenario 3.

4.3.3 THE FITNESS FUNCTION

The fitness function returns the strength of a candidate solution by which the nomination of any individual for next generation is depending on (Mawdesley et al, 2002). Fitness function measures the chromosomes' fitness in the population. Selection of a new population during the reproduction process greatly depends on chromosome's fitness value (Davis, 1991).

In this research, as the aim is to test the applicability of the proposed approach of station points to optimise 2D highway alignment, therefore, the critical cost components, as described in the previous sections, were combined linearly to form the total cost (C_{Total}), and the aim of the process was therefore to minimise;

$$C_{Total} = a_1 . C_{Length} + a_2 . C_{Location} \tag{4.6}$$

Where a_1 and a_2 are weighting factors of the cost components.

This fitness function was formulated so that the optimum alignment solution is compromised between the length and location costs. The solution should possess possible minimal length and skirts around the high cost field areas.

4.3.4 GA FORMULATIONS

4.3.4.1 Chromosome Representation

The principles of station points require each station point location to be defined by their X and Y coordinates for 2D highway alignments and X, Y, and Z coordinates for 3D alignments. These station points, as mentioned earlier, are treated as the decision variables of the problem.

Using the proposed design principles of this study, a candidate solution was represented using a number of station points. The total number of station points (N) was affected by different factors as listed below:

1. The length of the alignment. The longer an alignment, the more points may be required at any given precision. This criterion is affected by the land use features and the topography of the area.
2. The curvatures required to produce a smooth alignment. The more circuitous the alignment is, the more station points are required.
3. The precision of the study region that is defined by the size of the grid cells (refer to Section 4.3.1 above). Each grid cell holds an average value and represents the whole cell. When a better precision is required to represent a study area, smaller grid cell sizes would be used and consequently more station points are likely to be necessary for better alignment representation.

Using a smaller or greater number of station points than required is not recommended. A smaller number may not always be enough to configure the solution and would perhaps lead to production of a rough solution. A greater number may overburden the computation time required to process the search and may cause some disturbances. Therefore, the number of the station points was linked with the length of the alignment and the grid cell size. The following relations were suggested to determine the minimum and maximum limits for the station point number determination:

$$N_{Lower} = \frac{L_s}{Grid\ Cell\ width\ x\ 2} \tag{4.7.1}$$

$$N_{Upper} = \frac{L_s\ x\ 1.5}{Grid\ Cell\ width} \tag{4.7.2}$$

$$N = \left(\frac{N_{Lower} + N_{Upper}}{2}\right) \pm\ x \tag{4.7.3}$$

Where N_{Lower} and N_{Upper} are the limits of the lowest and highest number of station points that may be used for the alignment representation and L_s is the straight distance between the two termini points. *N* is an average rate of the number of the station points between the two limits, and *x* is a user defined number.

For instance, a simple area with almost low variations in the land features and few obstructions where alignments tend to be straighter than circuitous, an *N* number closer to the limit of N_{Lower} is preferable and vice versa to select a number closer to N_{Upper}.

Using the ideas of station points introduced above, the method defined alignment through generating N station points along the centreline of the alignment. The station points were considered as the decision variables of the alignment. Each decision variable (station point) in the chromosome was then considered as a gene encoded to X and Y coordinates. Starting from one of the termini points at one end (start point) towards the other one (end point), which were considered as known in this study, the station points were interconnected sequentially according to the order that they possess which were attached to them during the initialisation. Thus, a candidate solution was formed.

The chromosome map was as shown in Figure 4.7, in which it typically represents the alignment shown in Figure 4.8. The chromosome map contains the two-dimensional coordinates of each station point. The station points appear in the order in which they occur along the length of the alignment. The order of the station points were fixed and kept constant throughout the successive generations until end of the search process. The station points' order of the final alignment solution would have been the same as the order of the initial candidate when it was first generated but with a different configuration. This means that the search process tries to manipulate the station points' position and find the configuration that yields an optimum 2D highway alignment solution.

Index (i)	0	1	2	…	i	…	…	N-1
Individual (j)	X_0	X_1	X_2		X_i			X_{n-1}
	Y_0	Y_1	Y_2		Y_i			Y_{n-1}

Figure 4.7: Genetic representation of a chromosome for a 2D highway alignment

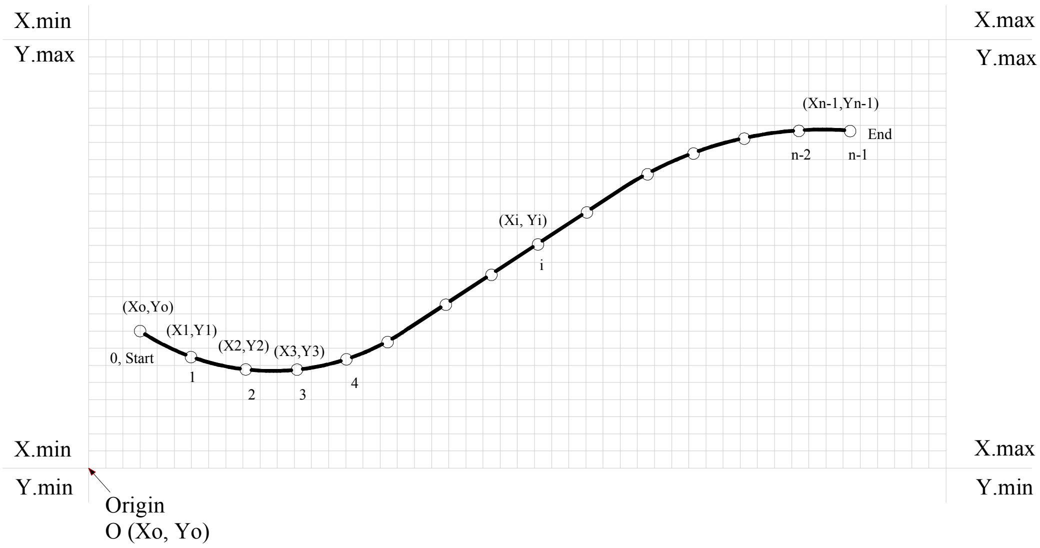


Figure 4.8: Graphical representation of a chromosome for a 2D highway alignment

4.3.4.2 Initial Population Generation

A random initial population of size (P_{size}) was generated such that:

- All station points were within the study area so that each gene of a chromosome was assigned random X and Y coordinate values, as follows.

$$X_{min} \le X_i \le X_{max} \quad (4.8.1)$$

$$and\ Y_{min} \le Y_i \le Y_{max}, \quad (4.8.2)$$

Where

i is the index or series of the station points in the order they occur in the solution chromosome; $i = 0,1, 2, \dots , N-1$, and

X_{min} and X_{max} are the lower and upper limits of the study area in X direction, and Y_{min} and Y_{max} are the lower and upper limits in Y direction.

Note: if the total number of the station points is equal to N including the two termini points, then considering 0 as the first terminal point would make the second terminal point N-1.

- The 2D components of each gene were encoded using floating point numbers.
- The first and last points (0 and N-1) were fixed as the required terminal points.
- The station points were sorted in the order of their X values ($X_i \le X_{i+1}$). This process was specific to the initial population generation only and it was done just to produce initial candidates with less windings.

An initial solution that is randomly generated in a 2D area using the above procedures may look like the one shown in Figure 4.9.

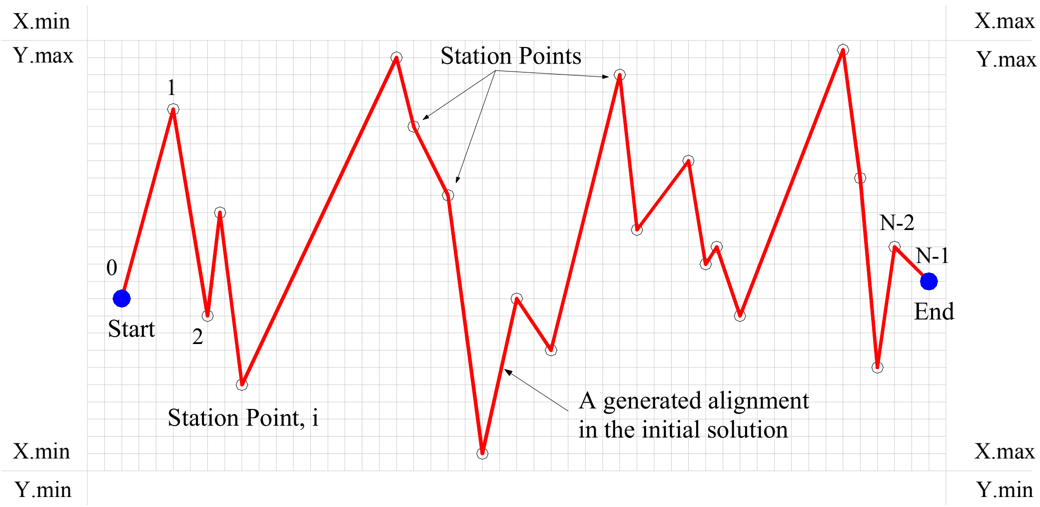


Figure 4.9: A randomly generated possible solution in the initial population

4.3.4.3 Selection

The natural selection and survival of the fittest concept allows the good individuals at the top of a ranked population to survive for mating while discarding the individuals at the bottom of the population (Haupt & Haupt, 2004). It is recommended that parents are randomly selected using a scheme which favours the more fit individuals (elitism) to breed in the next generation (Coley, 1999)(Beasley et al, 1993). Therefore, most of the selection methods are designed to satisfy the selection of the elite chromosomes. The selection methods like roulette wheel (fitness-proportionate selection), tournament, ranked, and genitor (steady state) select individuals based on either individual's proportionate fitness to the average population fitness or based on the proportionate rank of the individual in a ranked population (Haupt & Haupt, 2004)(Mitchell, 1996)(Lee, 2003)(Baker, 1985)(Mawdesley et al, 2002)(Chakraborty, R.C., 2010)(Goldberg & Deb, 1991)(Blickle & Thiele, 1995). Other selection methods like Boltzmann and generation gap are also available.

The design of the selection operators of this study allows the whole population individuals participate breeding so that the entire spaces of the search area are explored. The design also does not allow any individual, including the fittest one, to be selected twice as is the case with many selection schemes (Beasley et al, 1993). In the model for 2D highway alignment, two different selection schemes were developed and tested. The mechanisms of these two schemes were as follows:

i) Ranking – Based Selection Scheme (RnkSS)

In this study the entire population was selected to undergo crossover starting from the first individual until the last candidate in the population. Using the RnkSS scheme, the couples were sequentially selected depending on their rank (fitness) in a sorted population. This selection method is also known as an incest mating method due to the similarities of some properties of the selected couples according to their rank in the population (Eshelman & Schaffer, 1991). This operator gives higher chances to the top ranked individuals to produce better offspring than the bottom ranked ones. However, the role of the other operators (crossover and mutation) remains valid.

Figure 4.10 below shows the algorithm that processes the RnkSS.

<u>RnkSS Algorithm</u>

Note: (Zero) was considered for counting the individuals in the population to match a zero based array performance in the programming language.

```
For i = 0 to ((P_size -1) - 1)   step 2
        Select Chr (i)
        Select Chr (i+1)
        Perform crossover
Next i
```

Figure 4.10: The procedure for RnkSS

Let P_{size} be the population size and Chr denote a chromosome in the population.

The sequential selection of the parents, which may share similar information at some stages of the process, could lead to produce offspring with almost similar characteristics as their parents. This happens when the selected individuals undergo crossover with no mutation because of the mutation probability. These identical solutions, when they are sorted, most probably take positions around their parents. Having this phenomenon, the entire population would quickly be dominated by the offspring resulting from the top parents (best ranked individual whose positions are at the top of the population) and end up with a premature convergence while leaving the chance for mutation only to perform slow and new random searches to produce better candidates. This could cause the solution stuck at local optima at early generations before the search explore the entire space. To overcome such problems, the design of the crossover and mutation operators must guarantee keeping the genetic information as diverse as possible and avoiding early convergence.

To reduce the risk of this selection method, another method, which is called the Reverse Selection Scheme (RevSS), was suggested. Beasley et al (1993) agreed that a well designed selection method may reduce the risk of premature convergence. This is due to the fact that the application of GA is theoretically based on an infinite population size rather than a finite number.

ii) Reverse Selection Scheme (RevSS)

This method was designed to make the contribution of the entire population more effective and to help discover the genetic information that is hidden by the high fitness valued individuals. The RnkSS mentioned earlier, gives higher chances for the good individuals at the top of the population to produce better solutions than those parents ranked at the bottom of the population. RevSS was designed as an alternative to RnkSS to give equal chances to the best and worst individuals and selecting them as a two couple to undergo crossover. This was made under the assumption that the individuals with high fitness value may also handle some useful information about the optimum solution.

This selection scheme selects the best parent (rank 0) to breed with the last one (rank P_{size}-1) and the second rank individual to breed with the next (rank P_{size}-2), and so on. Figure 4.11 below shows the algorithm that performs the RevSS.

RevSS Algorithm

Note: (Zero) was considered for counting the individuals in the population to match zero-based array performance in the programming language.

```
For i = 0 to (Psize/2 - 1)   step 1

Select Chr (i)

Select Chr (Psize -1-i)   {(Psize - 1) is equal to the last individual
because of zero-based array in the programming language}

Perform crossover

Next i
```

Figure 4.11: The procedure for RevSS

Let P_{size} be the population size and Chr denotes a chromosome in the population.

4.3.4.4 Crossover Operator

After the selection is made, the chromosomes undergo crossover. Crossover is the process of exchanging information between two selected parents to produce offspring with mixed genes (Mawdesley et al, 2002). Crossover alters the chromosomes during reproduction and participates in the evolution (Coley, 1999).

The different crossover methods were tested in this study were as follows:

i) Single Point Crossover (SPCrO)

This operator is the simplest crossover form of standard genetic algorithm. It is common with binary representation. With this method a station point position within the chromosome length of the parent chromosomes was randomly determined, let it be *p*, and the points located at the right of this position were swapped with the same set of points of the other parent candidate (Note: this operator can also be used with fixed points) (Chakraborty, 2010) (Haupt & Haupt, 2004). Figure 4.12 to Figure 4.14 illustrate the mechanisms of this crossover.

Parent A				← *p*; Crossover point				
Index (i)	0	1	2	…	i	…	…	n
Individual (A)	X_0	X_1	X_2		X_i			X_n
	Y_0	Y_1	Y_2		Y_i			Y_n

Parent B				← *p*; Crossover point				
Index (i)	0	1	2	…	i	…	…	n
Parent Individual (B)	X_0	X_1	X_2		X_i			X_n
	Y_0	Y_1	Y_2		Y_i			Y_n

Figure 4.12: Parent chromosomes A and B before one point crossover

Offspring A^-				← swap point between the two parts				
Index (i)	0	1	2	…	i	…	…	n
Offspring (A^-)	X_0	X_1	X_2		X_i			X_n
	Y_0	Y_1	Y_2		Y_i			Y_n

Offspring B^-				← swap point between the two parts				
Index (i)	0	1	2	…	i	…	…	N
Offspring (B^-)	X_0	X_1	X_2		X_i			X_n
	Y_0	Y_1	Y_2		Y_i			Y_n

Figure 4.13: Offspring after one point crossover

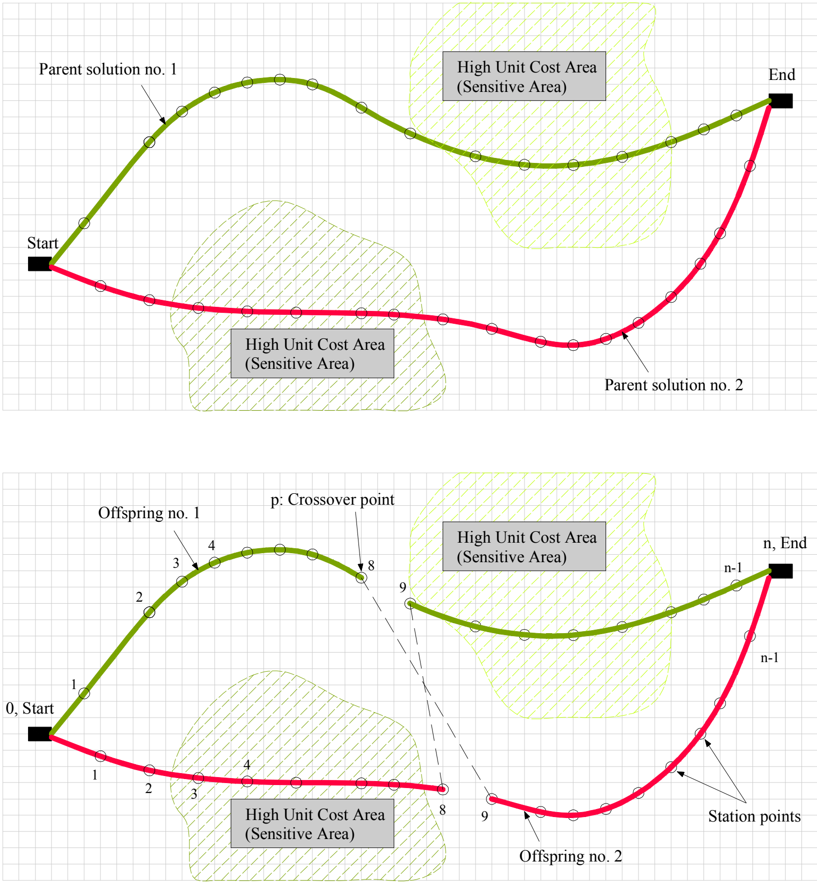


Figure 4.14: Graphical representation of one point crossover on the parent alignment solutions

One of the disadvantages of this method is that, when an alignment is long with a relatively big number of station points, the operator will produce offspring from a combination of two large parts. Swapping large parts may lead to a poor exploration and inefficient search. Therefore, double point crossover was also suggested.

ii) Double Point Crossover (DPCrO)

This method is similar to the SPCrO above but, instead of generating one point, it generates two random positions, let them be p_1 and p_2 provided that $p_1 < p_2$, and swaps the genetic information between them. This operator was also tested with using fixed point positions.

iii) Triple Point Crossover (TPCrO)

This operator generates three random locations along the selected parent chromosomes to perform crossover. Let these points be p_1, p_2, and p_3 provided that $p_1 < p_2 < p_3$. This operator swaps the genetic information between the start point and p_1 and between p_2 and p_3.

iv) Random-Multiple Point Crossover (RMPCrO)

As explained earlier, the configuration of the alignment with station points is complex, especially because the points are dependent on each other and therefore dealing with single or even double points may not always assist to produce good solutions. Moreover, when the environment of a study region is complex because of different land features, an optimum alignment may need to wind around the high cost areas thus requiring a circuitous alignment configuration. RMPCrO was then suggested to assist exploring wider areas and test various configurations. This crossover method is almost similar to uniform crossover approach of binary representation (Beasley et al, 1993). It randomly exchanges a number of genes, as alignment segments, in one exchanging task between any two randomly selected loci on the two parent chromosome.

This operator generates a random number (different numbers of crossover points) per each selected couple for crossover. This is to ensure different crossover structure and to keep the diversity of the genetic information pool as wide as possible. Let this number be R, between 1 and Cr; so that:

$$R \in \{1, C\} \tag{4.9}$$

Where; R is a random integer number that defines the number of crossover points and Cr is a user-defined parameter that defines the maximum crossover point numbers.

Under any circumstances Cr cannot be greater than or equal to N (Cr<N, where N is the total number of the stations points). Next, R locations are then generated between 1 and N provided that the generated locations are sorted such that $R_1 < R_2 < R_3 < R_4 < \dots R_n$ and so on. These locations are determined on the selected candidate parents and the genetic information is exchanged.

A valid Cr as a maximum crossover point needs to be investigated. A large Cr may deteriorate the search and produce no distinct difference between the offspring and the parents, whereas a small number (single or double) might not be able to produce some necessary changes to avoid getting stuck at a particular alignment configuration. The implications of each case are discussed later in the experimental sections.

4.3.4.5 Mutation Operator

Mutation, as one of GA operators, works on a single child. It randomly alters one or more genes to restore or improve the genetic information. (Michalewicz, 1999) (Beasley et al, 1993).

Mutation, as described by Holland, prevents the loss of diversity at a given bit position. So it provides insurance policy for the new children from being remained immutable at any particular locus (Mitchell, 1996).

Mutation application varies from a problem to another one. They often devised according to the nature of the problem. Specially designed mutation called problem-specific mutation operator (Jong and Schonfeld, 2003). The design of a good mutation is affected by the nature of the problem and the way that the solutions are represented.

In this study several mutation methods were suggested and tested. The following section gives the detail.

i) Standard GA Mutation Operator (Uniform Mutation; SM)

This method is a common operator with binary representation. It randomly selects a gene position and flips over its code; e.g., if the selected gene is coded as (0) it is turned to (1) and if it is (1) it is flipped to (0) (Forrest, 1996) (Hrstka & Kucerova, 2004).

The same principle was extended to be used with real or floating number representation methods. A position, say *p*, was randomly generated provided that (1 ≤ *p* ≤ N). The decoded gene values (X and Y coordinates) at this position were randomly replaced with new values within the limits of the boundary area.

Let X_p and Y_p be the coordinate values at *p* position before the standard mutation method is applied. The operator then assigns new values as:

$$X_p^- = X_{min} + rnd\,\{X_{max} - X_{min}\} \quad (4.10.1)$$

$$Y_p^- = Y_{min} + rnd\,\{Y_{max} - Y_{min}\} \quad (4.10.2)$$

In this study a probability rate parameter called Point Mutation Rate (PMR) was associated with this mutation method. PMR is the selection rate of the number of station points for mutation per each selected individual. It was expected that SM behaves differently with different PMR values. Therefore, a PMR value needed to be tuned based on the search performance. Figure 4.15 explains how PMR was used to select points for mutation.

For each station point, s = 0 to N-1
Generate an integer number randomly between (1 and 100); let it be A
If A ≤ PMR: Mutate the station point (s)
If A > PMR: Next s (select another station point)
Continue the process until N-1 is reached

Figure 4.15: An explanation for PMR application

This method, as an average, guarantees the selection of {Int (PMR*N/100)} station points. It should be noted that the individuals that underwent mutation were selected using Individual Mutation Rate probability (IMR). Thus, the number of individuals were mutated was equal to {Int(IMR*P_{size}/100)} and they were randomly selected.

ii) Grouped Point Mutation (GPM)

With this method a group of points, those linked sequentially, are moved. The method was designed to deal with segments of alignment (made of several station points) instead of single points. The selection of a group of points and moving them together was to make a big jump or wide move in one mutation step. This mutation can also be called 'Segment Mutation' as it deals with a segment of alignment.

This operator, in addition to its ability of moving segments or parts of the candidate solutions, also straightens some parts of the alignments. The trend of this mutation was therefore to enhance the search, introduce straight segments, and produce smoother solutions.

GPM, as with SM, selects a gene position R randomly and assigns new X and Y coordinate values randomly. This method associates a number of sequential station points with the mutated station point at R through the generation of two more locations (l_1 and l_2) provided that $l_1 < R < l_2$. Then, all the genes (station points) that locate between l_1 and R on the left side of R and l_2 and R on the other side are reallocated and put on a straight line connecting the newly generated gene at R with the selected genes at l_1 and l_2.

The mathematical steps of this mutation are detailed below.

1. Let A be a selected offspring chromosome before applying GPM or Segment Mutation.
2. Let R be a randomly generated position along the chromosome of length *N*.

$$R = rnd\,\{1, N-1\} \tag{4.11.1}$$

The selected gene at position R is decoded to X and Y coordinate values.

3. Assign new values to X and Y coordinates at R:

$$X_R^- = X_{min} + rnd\{X_{max} - X_{min}\}; \quad where\ X_{min} \leq X_R^- \leq X_{max} \tag{4.11.2}$$

$$Y_R^- = Y_{min} + rnd\{Y_{max} - Y_{min}\}; \quad where\ Y_{min} \leq Y_R^- \leq Y_{max} \tag{4.11.3}$$

4. Generate two more random positions l_1 and l_2:

$$l_1 = rnd\ \{1, R - 1\} \tag{4.11.4}$$

$$l_2 = rnd\ \{R + 1, N - 1\} \tag{4.11.5}$$

5. Alter the location coordinate values of the points located between l_1 and R:

$$X_i = X^-_R - \left[\frac{X^-_R - X_{l1}}{R - l_1}\right] x\ (R - i) \qquad for\ all\ i = (l_1 + 1), \dots, (R - 1) \tag{4.11.6}$$

$$Y_i = Y^-_R - \left[\frac{Y^-_R - Y_{l1}}{R - l_1}\right] x\ (R - i) \qquad for\ all\ i = (l_1 + 1), \dots, (R - 1) \tag{4.11.7}$$

6. Alter the location coordinate values of the points located between R and l_2 :

$$X_i = X^-_R - \left[\frac{X^-_R - X_{l2}}{l_2 - R}\right] x\ (i - R) \qquad for\ all\ i = (R + 1), \dots, (l_2 - 1) \tag{4.11.8}$$

$$Y_i = Y^-_R - \left[\frac{Y^-_R - Y_{l2}}{l_2 - R}\right] x\ (i - R) \qquad for\ all\ i = (R + 1), \dots, (l_2 - 1) \tag{4.11.9}$$

Figure 4.16 to Figure 4.18 show a mathematical example of this mutation.

Index	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	…	…	*N*
chromosome A	X_s	45	58	75	38	135	75	84	112	78	144	152	…	…	X_e
	Y_s	16	44	33	57	24	18	42	22	15	36	2	…	…	Y_e

Figure 4.16: An offspring before GPM (Step 1 above)

Index	*0*	*1*	*2*	*l_1*	*4*	*R*	*6*	*7*	*8*	*l_2*	*10*	*11*	…	…	*N*
Chromosome A	X_s	45	58	75	38	86	75	84	112	78	144	152	…	…	X_e
	Y_s	16	44	33	57	30	18	42	22	15	36	2	…	…	Y_e

Figure 4.17: A chromosome with positions l_1, l_2, and R using GPM (Step 2, 3, and 4 above)

Index	*0*	*1*	*2*	*l_1*	*4*	*R*	*6*	*7*	*8*	*l_2*	*10*	*11*	…	…	*n*
Parent 1	X_s	45	58	75	80.5	86	84	82	80	78	144	152	…	…	X_e
	Y_s	16	44	33	31.5	30	26.2	22.5	18.7	15	36	2	…	…	Y_e

Figure 4.18: An offspring after GPM application (Steps 3, 5, and 6 above)

The above mathematical representations are graphically shown in Figure 4.19 and Figure 4.20. The new offspring possesses less length, is less likely affected by the high unit cost area, is more tidy, and has straighter segments. Thus, the new offspring is more likely to be fitter than the parents.

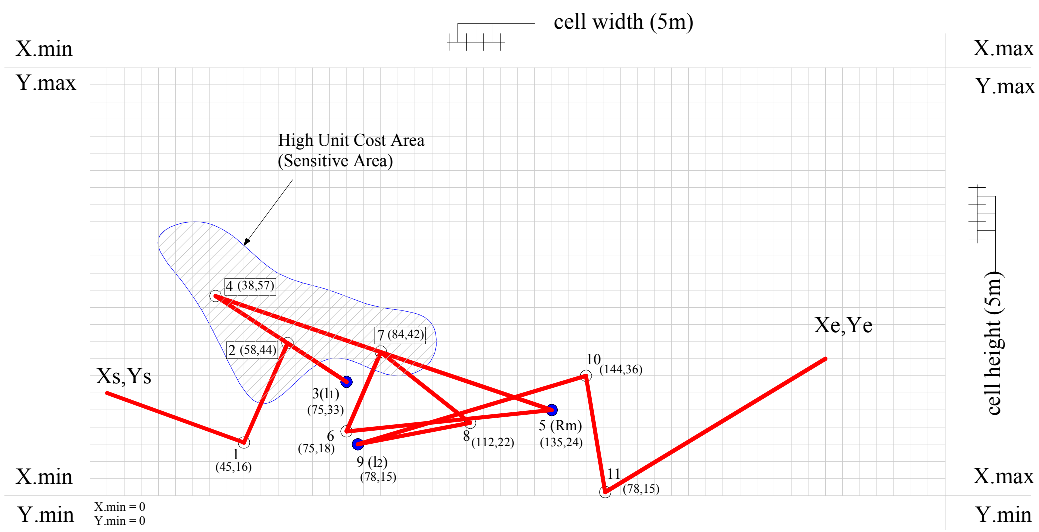


Figure 4.19: A selected candidate before applying GPM

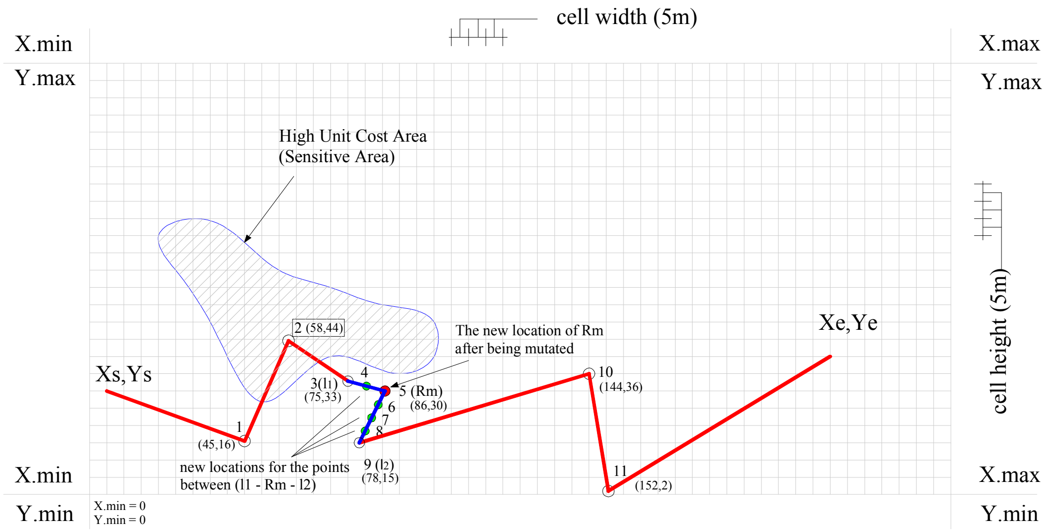


Figure 4.20: An offspring after applying GPM

In this study and for 2D highway alignment optimisation, GPM was applied using two different strategies. These strategies handle the boundary of the study area for each selected gene in two different ways, as follows.

1. Random Simple Area consideration (RSA): This method applies simple random mutation to the coordinate values at R within the boundary limitation of the study area as described by equations 4.11.2 and 4.11.3. This method allows free searches within the boundary area and lets any bit of the area be explored at any time.
2. Random Dynamic Area consideration (RDA): This method randomly reduces the search spaces in X and Y directions and applies mutation within it. For each selected individual to undergo mutation, this operator generates a random dynamic area factor R_{area} so that $R_{area} \in \{0, \ldots, 1\}$ and is applied as follows.

- Randomly generate a single precision value between 0 and 1 (R_{area}) and then find the limits of X and Y, as below:

$$X_{upper\ limit} = X_R + (X_{max} - X_R) * R_{area} \qquad (4.12.1)$$

$$X_{lower\ limit} = X_R - (X_R - X_{min}) * R_{area} \qquad (4.12.2)$$

- Randomly mutate the selected gene at R within the new defined limit:

$$X_R^{-} = X_{lower\ limit} + Rnd\ \{X_{upper\ limit} - X_{lower\ limit}\} \qquad (4.12.3)$$

These procedures can similarly be repeated for Y coordinates, too.

It should be mentioned that after the mutation method was applied, the offspring candidates were sent for evaluation and the fitness value of each was calculated. The two populations (the parents and the offspring) were then combined to form one population of size ($2xP_{Size}$). The new double-sized population was sorted according to their fitness and a population of size (P_{Size}) from top individuals was selected to continue breeding in the coming generations.

4.3.4.6 Termination Criteria

In practice and depending on the nature of the problem it might not be easy to know when or where to stop the search process without knowing anything about the nature of the optimum solution and its fitness value. Changing a single parameter value during the experiments might bring different changes on the overall behaviour of the search process and the configuration of the final solution. Therefore, it will be wise to give some flexibility to the model during the search process before deciding on where, when, and how to stop the process.

In this study if the best individual in the population remains unimproved over a number of successive generations (g_T), which is defined by the user, then the process is terminated; otherwise, reaching up to the total generation number (G), which is also specified by the user and supposed to be big enough to handle the whole search process, will terminate the search. Whichever of these criteria is satisfied, the process will come to a halt.

4.4 HORIZONTAL HIGHWAY ALIGNMENT CONSTRAINTS

The existing (traditional) approach of geometric design for horizontal highway alignment deals with curvature as one of the elements of safety and comfortable driving condition. The approach provides horizontal curves for smooth transition between the different directions. The horizontal curvature is defined by degree of the curve and is a function of the curve radii. In addition to safety, curvature is required to maintain the design speed as well. To provide a safe transition on a horizontal alignment, the horizontal curvature or degree of the curve at any given point must not violate the allowable limits specified by the standard requirements. As this chapter aims to investigate the viability of using the proposed design approach of station points for highway alignment optimisation, curvature is therefore not detailed here. This chapter takes the boundary of the study area as the only constraint. Extensive coverage of horizontal curvature is presented in Chapter 6.

4.5 EXPERIMENTAL TESTS AND THE RESULTS

So far, building the structure of the model has been discussed. The model is now available and requires some input parameters to make it ready for implementation. The model, at this stage, needs to be tested in two different ways:

- a test to make sure that the math of the model is correctly formulated,
- a test to investigate the validity of the proposed approach of station points for highway alignment optimisation.

These tests may help to tune and find valid parameter values that make the performance of the model optimal for highway alignment optimisation. The performance and the efficiency of the model were associated with these parameters and their relationships. For this purpose a testing regime was also set so that:

- to tune the different parameter values that result in producing a ‘good’ solution

- to test and tune the parameters on different world cases for reliability.

The list of the proposed main parameters for investigation was as below.

1. Length of the chromosomes represented by the number of the station points.
2. Population size.
3. Selection methods.
4. Crossover methods.
5. Mutation methods.
6. Mutation probabilities. According to the nature that the chromosomes were represented, the mutation probability was divided into two categories.
   - A probability rate to select the individuals to undergo mutation. This was called Individual Mutation Rate (IMR).
   - A probability rate to select a range of points (a number of points) within the selected chromosome for mutation. This was called Point Mutation Rate (PMR). Both are thought to be influential.

Determination, adjusting, and tuning these parameters were a key for the model to generate a global or near global solution. So, careful consideration must have been given to analyse these parameters.

For this purpose several tests were set out, varying between model verification and parameter tuning tests. The details of these tests are given in the following sections.

4.5.1 TEST GROUP 1: MODEL VERIFICATION

The intension of this test was to verify and test the genetic algorithms' formulations and to verify the applicability of the proposed station point design approach (SPDA) for highway alignment optimisation.

Before beginning, the main expressions of the study need to be unified. Table 4.2 lists the expressions' abbreviations and symbols that are used within the whole of this thesis.

Table 4.2: Table of abbreviations for some parameter names

Abbreviation	Description
W	World case. This is followed by a number to refer to the World case number.
N	Number of the station points, which defines the length of the chromosome.
CrO_Type	Crossover type
MM	Mutation method
IMR%	Individual mutation rate (or probability of individual selection to undergo mutation)
PMR%	Point mutation rate (or probability of selecting the number of station points for mutation)
SS	Selection scheme or selection method
T	Termination criteria
RnkSS	Ranked selection scheme
RevSS	Reverse selection scheme
SPCrO	Single point crossover
RSPCrO	Random single point crossover
DPCrO	Double point crossover
RMPCrO	Random multiple point crossover
G	Total generation number
g	Current generation number
FFV	Fitness function value

The first test was conducted on a 2D plain world (W_1). The problem was to find the optimum solution that connects two known termini points. The absolute solution for this problem was made known and had to be a straight alignment. It is worth mentioning that a similar notion of model validation was suggested by Parker (1977). He described the possibility of constructing restricted situations with “obvious” solutions.

The initial parameters took part in this investigation are identified in Table 4.3.

Table 4.3: Initial test parameter set

W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS	T
W_1	10	500	RSPCrO	SM	50%	2/N = 20%	RnkSS	G = 2000

Figure 4.21 shows the first study area set up of W_1 with a known optimum solution. The area is completely plain with unique location costs. The aim of the problem was to connect the known start and end points by an alignment. This test was to enable us to verify the performance of the genetic operators' formulations and their ability to locate a good solution with the use of station points.

The first test run was launched with considering length (construction) and location dependent costs in the fitness function (see equation 4.6). To demonstrate the performance of the genetic algorithm operators, two parents and their offspring were selected after RSPCrO and presented in Figure 4.22. The two parents were selected at generation 5 where a RSPCrO was applied at station point number 4.

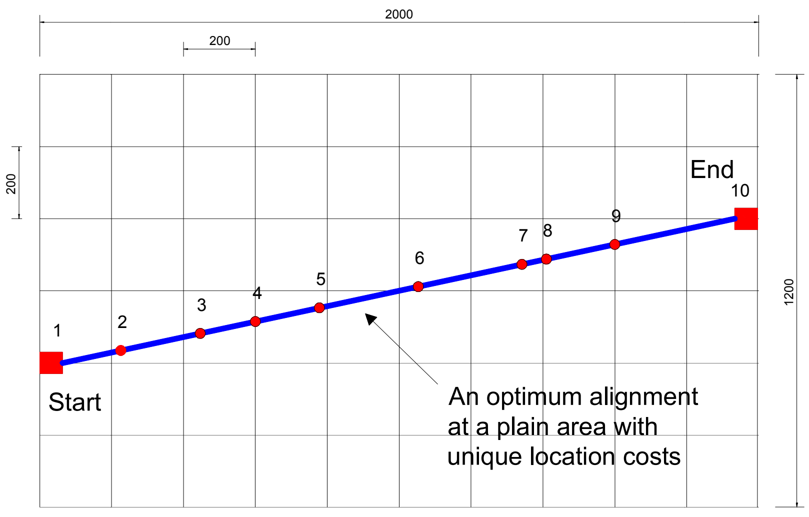


Figure 4.21: World W_1 set up with a known straight alignment solution

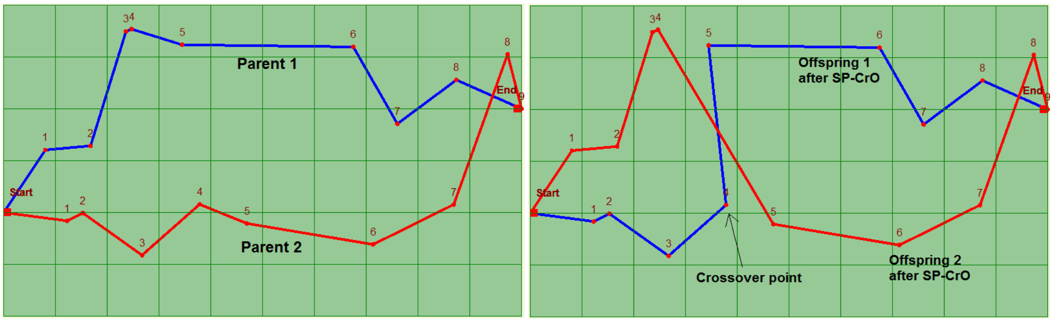


Figure 4.22: Left) Parents before RSPCrO; Right) Two offspring after RSPCrO at point 4

Figure 4.23 shows the resulting solution at generations 1,000 where 99.99% fitness of the known optimum solution was obtained. It should be noted that different station point configurations on the same straight line produce the same fitness values.

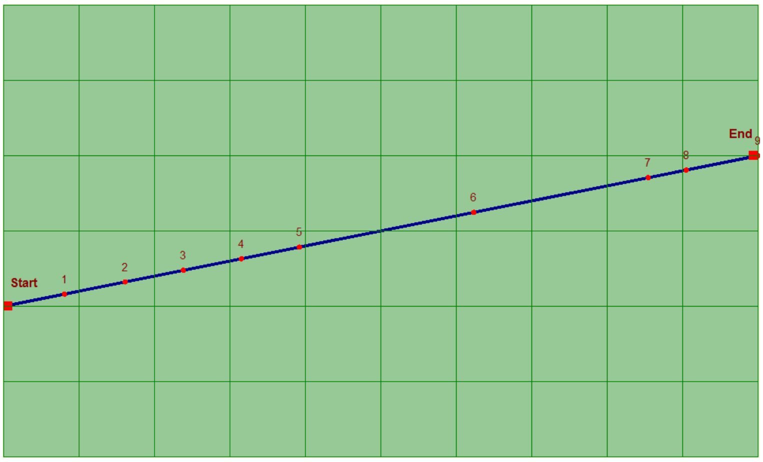


Figure 4.23: The best alignment solution at g = 1000

Another verification test was carried out using the same parameters, as in Table 4.3, except that the world was slightly changed where four cells at the middle of the region were assigned higher unit costs (W_2). This test was aimed at finding the ability of the search model to find alignment solutions that require making jumps over the high unit cost areas. The test results were good and promising, as shown in Figure 4.24.

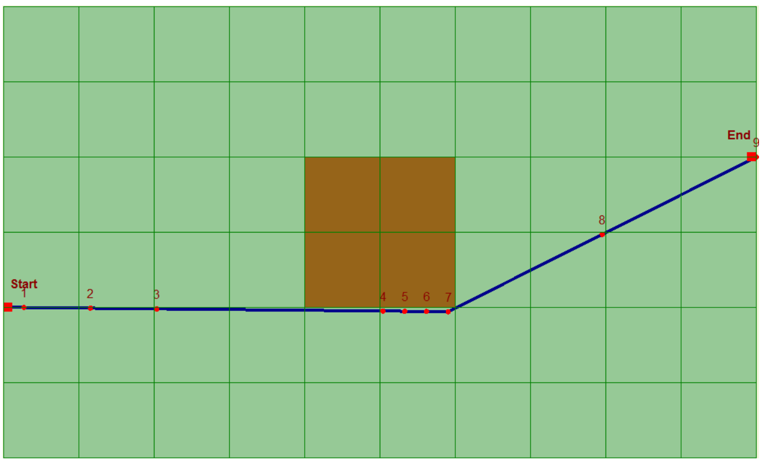


Figure 4.24: An optimum solution at g = 1000

Table 4.4 and Figure 4.25 compare the fitness results at different generations. They show that the simpler the study area, the quicker is the convergence and the fitter would be the optimum solution (e.g. compare the results at generation 25).

Table 4.4: The test result of the initial test at different generations

	W_1 results			W_2 results		
g	Alignment Length (m)	Fitness value (Unit Cost)	% of global optimum FFV	Alignment Length (m)	Fitness value (Unit Cost)	% of global optimum FFV
Known optimum solution	**2039.60**	**613,919.6**	**Optimum = 100%**	**2094.42**	**628328.15**	**Optimum = 100%**
0	2,629.37	791,439.24	71.08%	3,152.88	1,090,093.52	26.50%
25	2,117.91	637,491.25	96.16%	2,228.90	670,897.21	93.22%
50	2,044.89	615,511.98	99.74%	2,102.22	632,767.56	99.29%
100	2,039.72	613,956.96	99.99%	2,099.36	631,907.29	99.43%
500	2,039.61	613,922.85	99.99%	2,097.52	631,354.79	99.51%
1,000	2,039.61	613,922.13	99.99%	2,097.52	631,353.67	99.52%
2,000	2,039.61	613,922.04	99.99%	2,095.66	630,792.89	99.60%

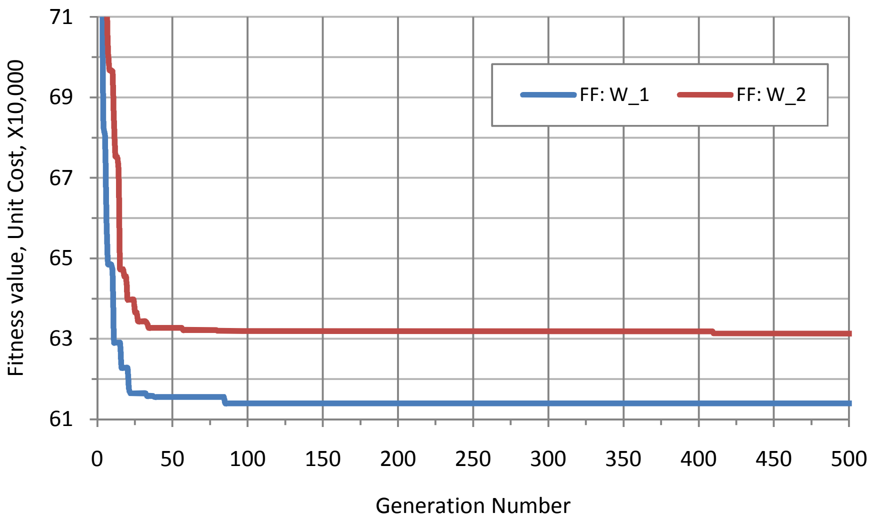


Figure 4.25: Fitness graph results of the two scenarios W_1 and W_2

Two more world cases were also developed with relatively big dimensions (12000m x 6000m). One of the worlds (W_3) was made plain, as with (W_1), and the other world case (W_4) was made with several high unit cost land features (500 unit cost/m^2). These worlds were created to test the implications of world sizes and different high unit cost locations on the model behaviour. The optimum solution of the world case W_3 at generation 1,000 is shown in Figure 4.26, which is optimum by 99.99% of the known optimum fitness value.

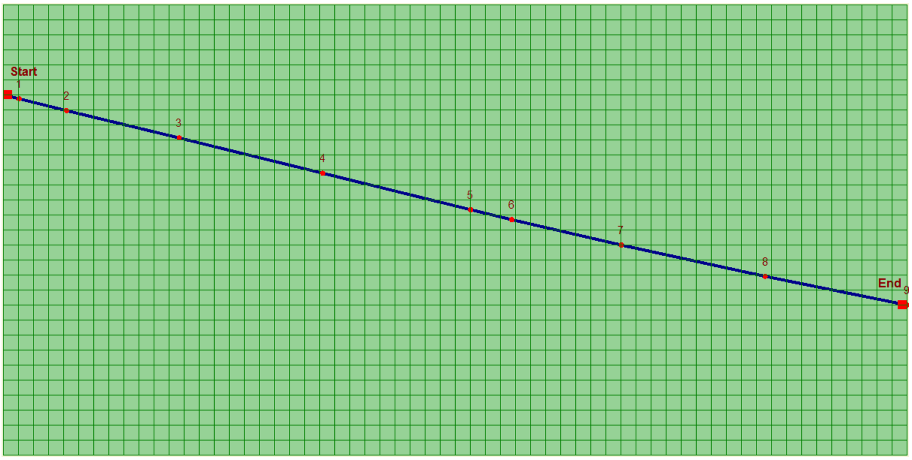


Figure 4.26: The optimum solution at generation 1000

The test results with W_4 showed that different optimum alignment locations with almost the same fitness value are available. The fitness value of the solution at generation 2,000, as shown in Figure 4.27, is 4,442,664 which is as 99.82% optimum as the known (absolute) global optimum solution (fitness: 4,434,634) where the latter skirts the boundary of the high cost areas.

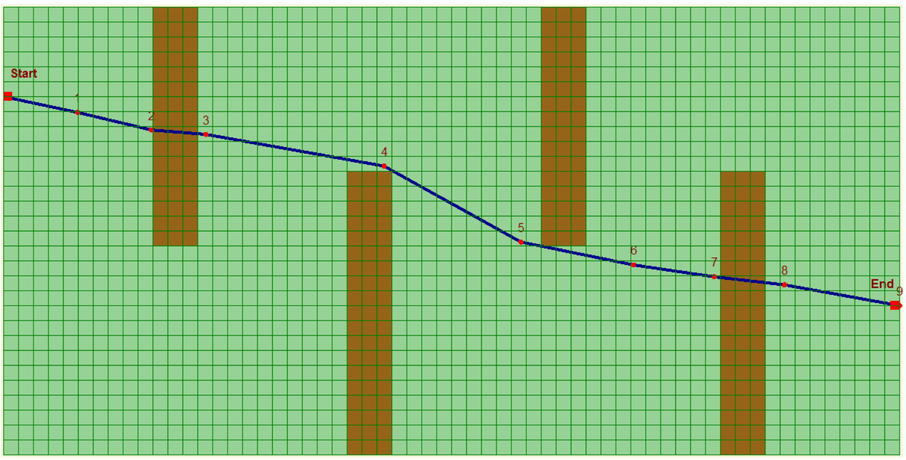


Figure 4.27: An optimum solution at generation 2000

It was noticed that the alignment crosses the high unit cost areas. Such cases are likely to happen when the fitness of the solution outside the high cost areas is almost similar or higher to that which crosses through these areas. If it is desired by the planner to avoid these high cost fields, the process needs to either assign a higher unit cost value to these cells or to use a weighting factor. Figure 4.28 shows an optimum solution with fitness (5,943,689) after increasing the location cost by a factor of 10. The solution represents (96.12%) of a known global solution with a fitness value of (5,722,109). The accuracy of the solution was due to the number of the points. This also proves that the model is sensitive against different cost factors. Table 4.5 presents the results of the different scenarios with W_3 and W_4. Figure 4.29 compares the convergences of W_3's and W_4's different cases.

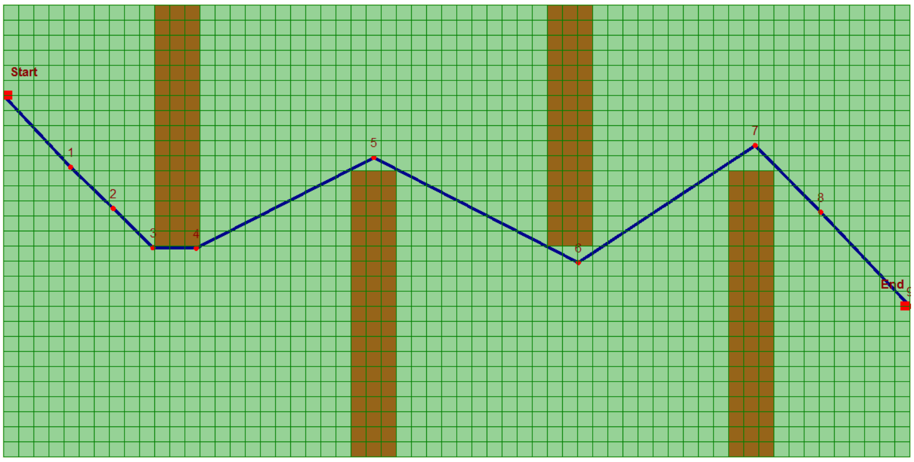


Figure 4.28: An optimum solution with weighting location cost

Table 4.5: The fitness results at different generations

	W_3 results		W_4 results (with no location weight factor)		W_4 results (with location weight factor = 10)	
g	Fitness value (Unit Cost)	% of global optimum FFV	Fitness value (Unit Cost)	% of global optimum FFV	Fitness value (Unit Cost)	% of global optimum FFV
Known optimum solution	**3,819,922**	**Optimum = 100%**	**4,434,634**	**Optimum = 100%**	**5,722,109**	**Optimum = 100%**
0	5,440,868	57.56%	6,496,576	53.5%	9,702,524	30.43%
25	3,869,667	98.69%	4,660,020	94.91%	7,499,226	68.94%
50	3,823,800	99.89%	4,563,378	97.09%	6,968,434	78.22%
100	3,820,781	99.97%	4,488,723	98.78%	6,658,039	83.64%
500	3,819,943	99.99%	4,444,305	99.78%	6,008,008	95.00%
1000	3,819,939	99.99%	4,443,470	99.80%	5,947,024	96.00%
2000	3,819,936	99.99%	4,442,664	99.82%	5,943,689	96.12%

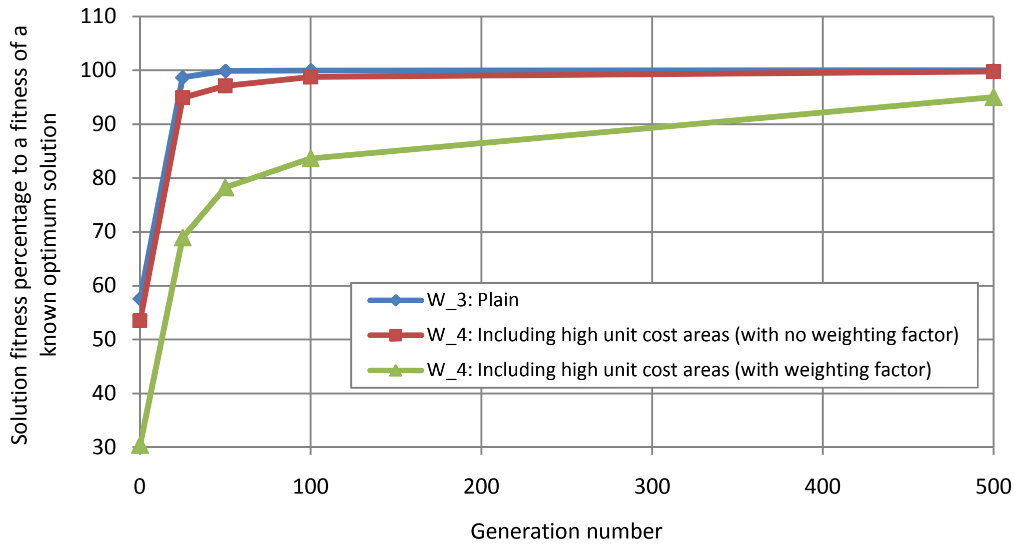


Figure 4.29: Comparisons of convergence rate of three different scenarios with W_3 and W_4

It was important to make sure that the model or the search algorithm can distinguish between different land features and different construction costs, especially when the model is under development, in order not to be confused by the characteristics of the problem parameters under uncertain conditions like those that have been mentioned.

The solution that has been presented in Figure 4.28 tells that the alignment can be made fitter if larger numbers of station points are used for better configuration. It can be seen that the alignment has bends that are configured by one station point (point locations 5, 6 and 7). This led to a rough alignment configuration and longer than the global optimum one. These imply that more station points would be necessary to make smoother changes, to produce shorter lengths, and consequently fitter solutions.

The above demonstration proved the correctness of the formulations and the potential of using GA to solve the highway alignment problem with the use of the station point approach. However, the presented problems are still simple compared with real world cases and therefore further formulations, developments, and tests on more complex world cases are required. The following sections try to address these cases in more detail.

4.5.2 TEST GROUP 2: TUNING THE MODEL PARAMETERS

This stage was dedicated to tuning the model parameters, listed in section 4.5 above, that help in producing optimum highway alignment in a relatively complex world.

For this purpose, another hypothetical world was manually developed (W_5) with a known global optimum solution. This world is shown in Figure 4.30 in which the optimum solution requires forward and backward bends (windings) in order to wind around the high cost areas. The blue shaded cells indicate high unit cost areas where the alignment must not cross. A relatively high unit cost equal to 10,000 unit cost/m^2 was considered for these locations (location cost) while the rest were assigned the value of 10/m^2. The construction cost for this scenario was set at 300 unit/m and the weighting factors of both components (the location and the construction costs) were kept equal and set to ‘1’.

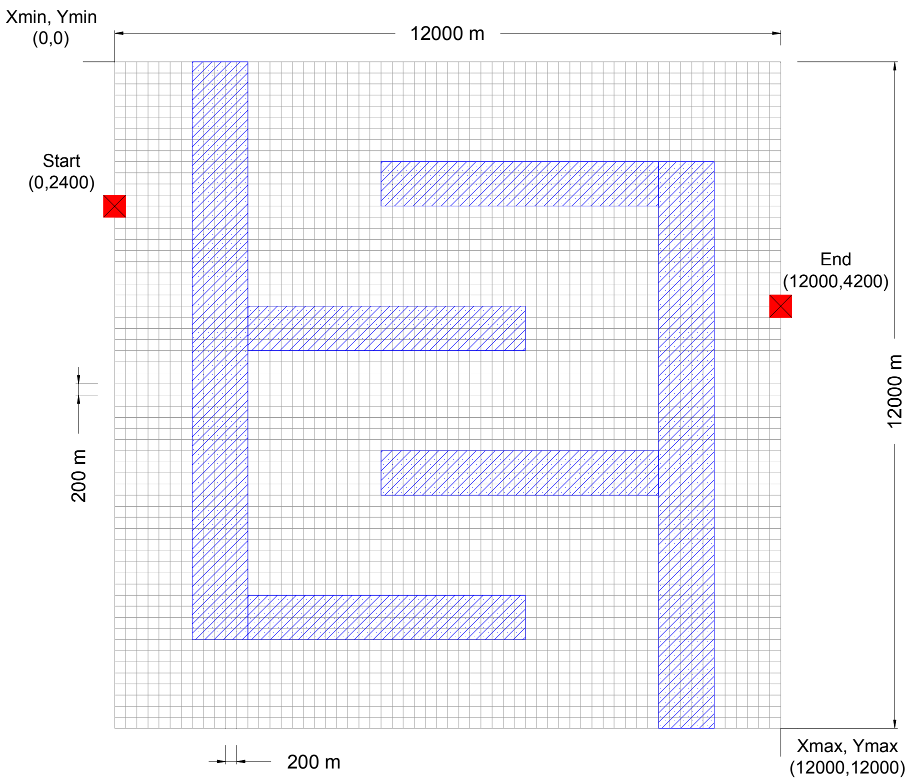


Figure 4.30: World W_5 set up

To understand the effect of each parameter on the model's search behaviour and the relation of each parameter to another, thousands of tests would be required. This would be a difficult task and requires a very extensive time. Therefore, only the parameters were thought influential and critical to the optimum solution were chosen for tuning whereby the decisions upon the solutions were made.

In this study, for a decision on a single parameter, 10 test runs as a minimum test number were chosen to be conducted to enable us understand the behaviour of the parameter in directing the search towards the optimum solution. The evaluations and judgements were then made based on the results obtained from the 10 test runs. The investigations and the observations are detailed in the following sections.

4.5.2.1 Number of Station Points

Two different world cases were used to investigate the implications of this parameter on the final alignment solution. With the first world case, almost the same experiments of section 4.5.1 were reproduced but this time, in addition to SM, GPM was also used. The aim was to tune, verify, and compare the two mutation methods with different numbers of station points. The parameters were first tested on the world W_4 using the test scenarios presented in Table 4.6.

Table 4.6: Initial parameter sets associated with GPM

Scenario	W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS
S1	W_4	60	1000	RSPCrO	SM	10%	20%	RnkSS
S2	W_4	60	1000	RSPCrO	GPM (RSA)	10%	Rnd – free	RnkSS
S3	W_4	10	1000	RSPCrO	SM	10%	20%	RnkSS

The results with GPM (S2), despite the use of a relatively large number of station points (N=60), showed a high potential of the mutation method in producing a global optimum solution. Figure 4.31 compares the performance difference between the SM with 10 station points and the GPM with 60 station points.

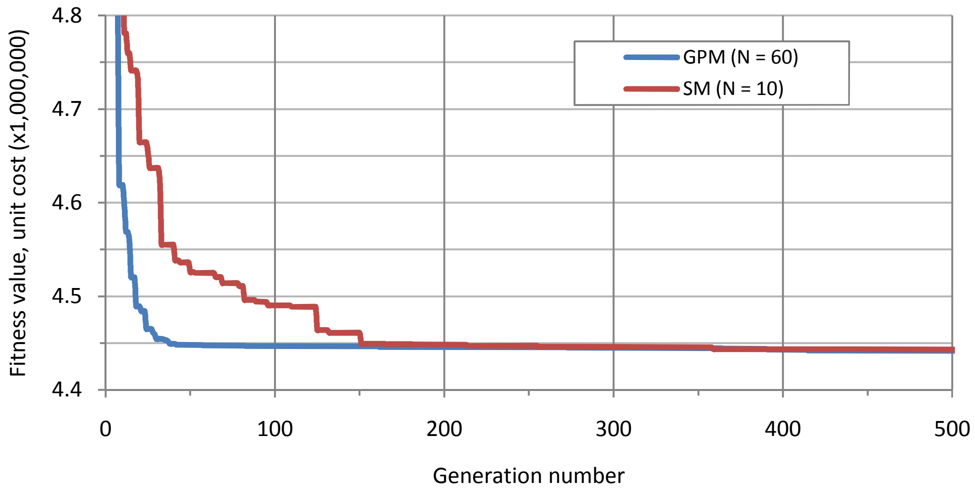


Figure 4.31: Comparison of fitness improvement between SM (N = 10) and GPM (N = 60)

The graph demonstrates that the GPM method can be considered as a good alternative successor for the simple mutation method SM. It shows that the convergence with GPM is quicker than the SM for obtaining nearly the same solution fitness despite the differences in the number of station points used to configure the alignment. It should be mentioned that the result of SM with 60 station points (S1) was not successful in producing a good solution. Figure 4.32 shows the difference between the two scenarios.

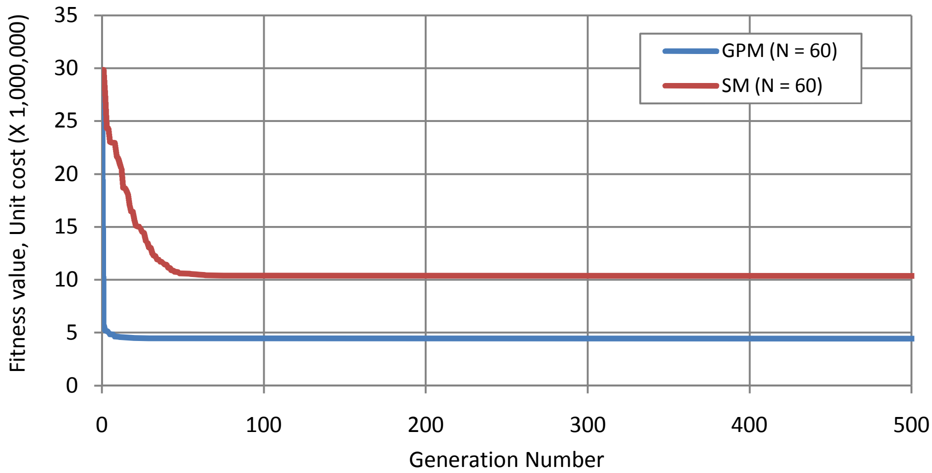


Figure 4.32: Performance comparison between SM (N = 60) and GPM (N = 60)

Figure 4.33 shows a solution result of GPM at generation 1000 (S2).

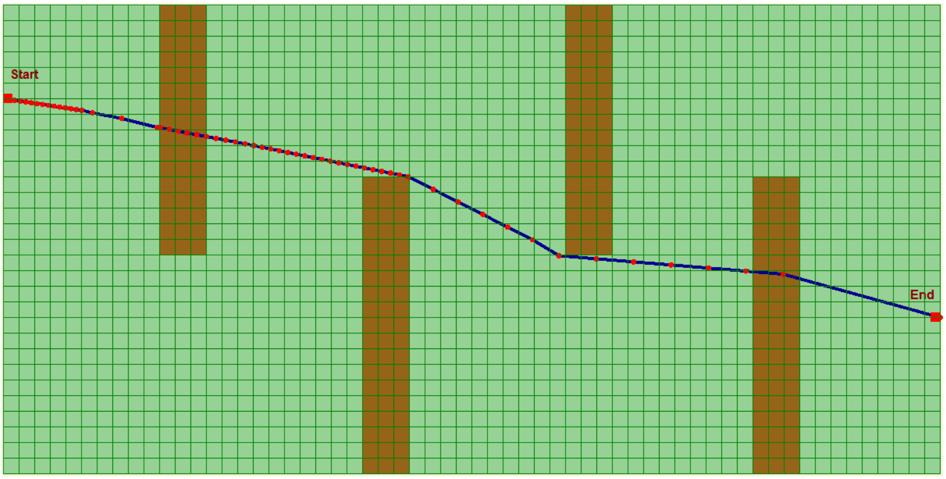


Figure 4.33: An optimum solution with GPM at generation 1000

The alignment solution, as shown above, was passed by a high unit cost. This happens, as mentioned before, when the cost or fitness ratio of a solution that passes by the high cost areas to the one that skirts these areas is too small or negligible or when the shorter one incurs less total costs due to less construction costs. The solution result of Figure 4.33 has a fitness value equal to 4,439,402 at generation 1000, which is optimum by 99.9% if compared with the 'absolute' optimum solution that skirts the high unit costs (the brown colours) with an optimum fitness value of 4,434,634.

To show the sensitivity of the model with GPM to different costs, a factor of 10 was given to the location cost and the result at generation 1000 of a population size 1000 was as in Figure 4.34.

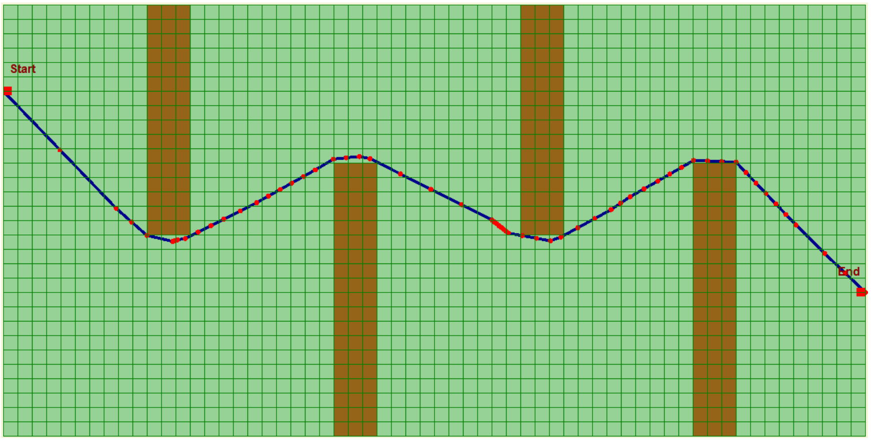


Figure 4.34: A solution with location cost weight of 10

Finding a good solution with world W_4 above looks to be straightforward and requires little effort (less than 50 generations with GPM) to find an optimum solution (see Figure 4.31). The case was totally different with the other world case (W_5). Getting a good solution was not as straightforward as with world W_4 and required tuning several parameters to yield an optimum solution.

Tests on W_5 started with the parameters shown in Table 4.7 and the test results showed that the model was unable to evolve the solution using these parameters and operators. A solution result is as shown in Figure 4.35.

Table 4.7: The initial parameter values for the tests on W_5

W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS	T
W_5	10	500	RSP-CrO	SM	50%	2/N = 20%	RnkSS	G = 2000

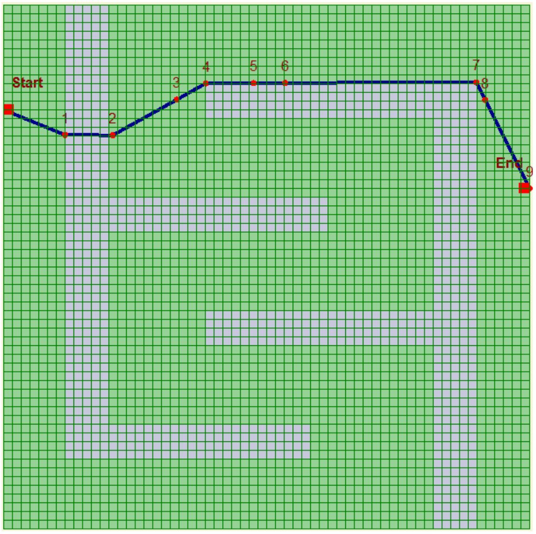


Figure 4.35: A non-evolved solution at generation 2000

It was concluded that highway alignments demands more points where it is proposed for relatively complex areas (worlds with different and irregular land features) to skirt around the different lands and, without this, the other parameters could not be judged. Therefore, a number of test scenarios were designed with up to 60 station points (based on equations 4.9.1 to 4.9.3) and the test results showed no noticeable improvement. This was because, in addition to the number of station points (30 to 60) which might have been enough to configure the required bends, the search may also depend on the other parameters, which also needed tuning. Therefore, the number of station points needed investigation in conjunction with other parameters. The following sections try to seek for an answer.

4.5.2.2 Various Population Size Tests

Delaver et al (2004) mentioned that the population size is problem dependent and needs to be exponentially increased with the complexity of the problem.

In this section, the effect of the number of individuals in the initial solution pool on the model behaviour and the search direction are investigated through the design of several test scenarios. Moreover, in conjunction with the population sizes, the performance of mutation methods of SM and GPM are also compared. Table 4.8 shows the different scenarios.

Table 4.8: Tests on various population sizes

Scenario	W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS
S1	W_5	30	50	RSP-CrO	GPM	10%	5%	RnkSS
S2	W_5	30	1000	RSP-CrO	GPM	10%	5%	RnkSS
S3	W_5	30	5000	RSP-CrO	GPM	10%	5%	RnkSS
S4	W_5	30	50	RSP-CrO	SM	10%	5%	RnkSS
S5	W_5	30	1000	RSP-CrO	SM	10%	5%	RnkSS
S6	W_5	30	5000	RSP-CrO	SM	10%	5%	RnkSS

The results showed that the higher the population sizes, the more stable is the performance of the model. Figure 4.36 shows that, per each mutation method, the number of solutions within 95%-100% of the most optimum solution is increased with the increase of the population size. Moreover, as shown by the fitness graph in Figure 4.37, the convergence with greater population sizes is quicker and the solution fitness is better. The results also show that GPM is more efficient than SM. These results are indications that the population size has a critical influence in providing genetic information and keeping the diversity of the population.

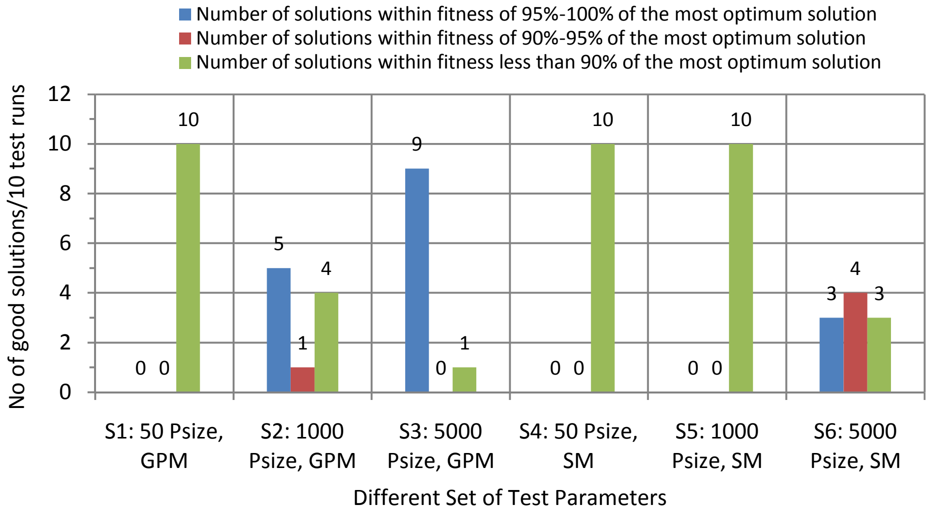


Figure 4.36: The effect of P_{size} and MM on the search stableness performance

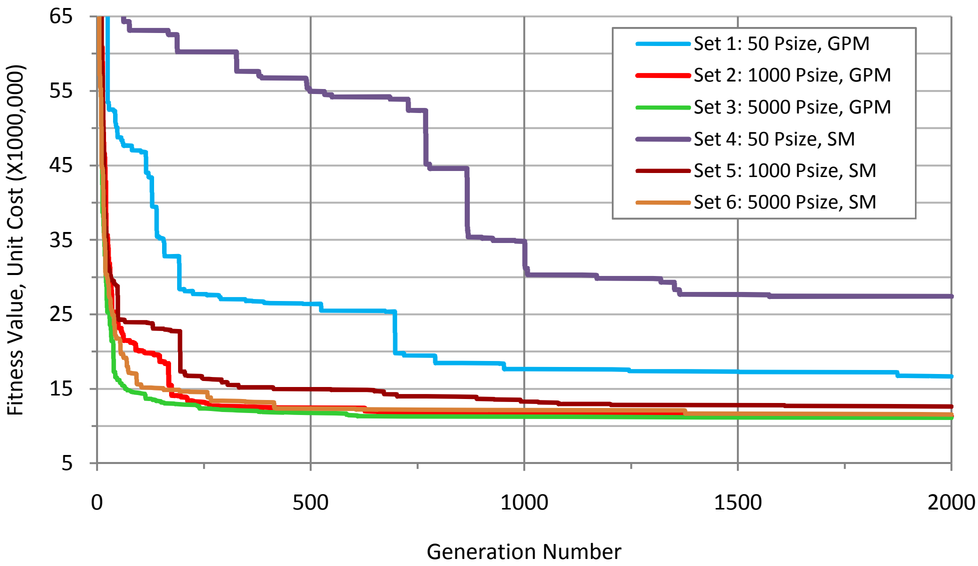


Figure 4.37: The effect of population size and mutation type on solution fitness and convergence

(The graph represents the best fitness obtained from the 10 test runs per each test scenario)

Figure 4.38 shows two best test results at generation 2000 for the parameters defined by S3 and S6, in which they possess length and fitness values equal to 35,850 m and 11,113,650 unit cost and 36,369 m and 11,274,346 unit cost, respectively.

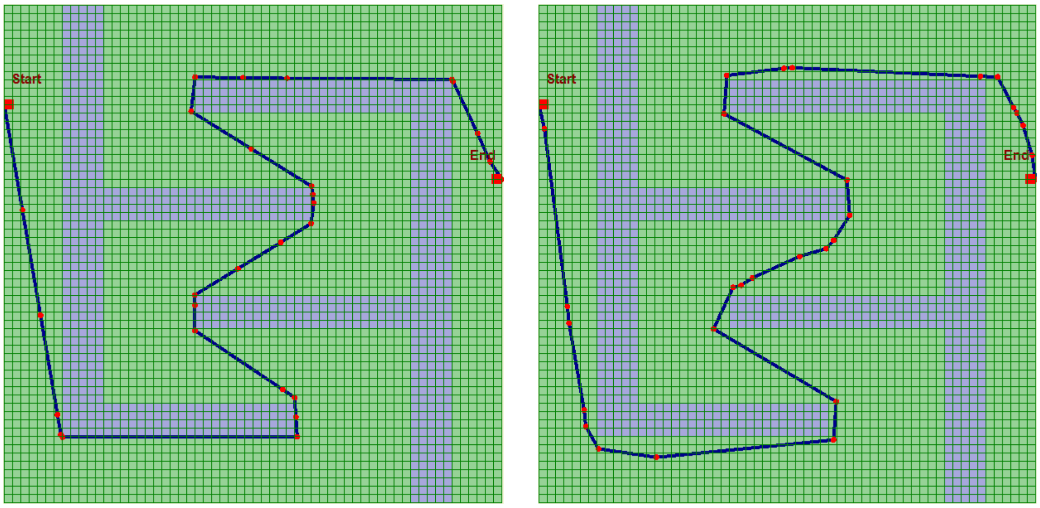


Figure 4.38: Left) a solution obtained from P_{size} 5000, at G 2000 and GPM 'S3', Right) a solution obtained from P_{size} 5000, G 2000 and SM 'S6'

It can be observed that, as in the fitness graph (Figure 4.37), the most improvement towards optimality is taking place within the first 250 generations and the remaining generations are used for fine tuning the solution.

4.5.2.3 Station Point Mutation Rate Tests (PMR%)

The implications of different PMR and mutation types were investigated using chromosome lengths of 30 and 60 station points. The test parameters of these tests were as in Table 4.9. It should be mentioned that the population size for these tests was set to 5000 based on the results from the previous section.

Recalling the previous explanations, PMR with SM selects a total number of station points for mutation equal to PMR*N/100 while with GPM the operator straightens a random number of station points up to PMR*N/100 on each side of the single mutated gene (refer to Sections 4.3.4.5 i and ii). Thus, the minimum PMR was set to 5% and 3% for the chromosome lengths 30 and 60 respectively so that the selection of at least one station point was guaranteed.

Table 4.9: Different sets of parameters for PMR tests

W	N	Psize	CrO_Type	MM	IMR%	PMR%	SS
Tests with SM							
W_5	30	5000	RSP-CrO	SM	10%	5%	RnkSS
W_5	30	5000	RSP-CrO	SM	10%	10%	RnkSS
W_5	30	5000	RSP-CrO	SM	10%	15%	RnkSS
W_5	30	5000	RSP-CrO	SM	10%	20%	RnkSS
W_5	60	5000	RSP-CrO	SM	10%	3%	RnkSS
W_5	60	5000	RSP-CrO	SM	10%	10%	RnkSS
Tests with GPM							
W_5	30	5000	RSP-CrO	GPM	10%	5%	RnkSS
W_5	60	5000	RSP-CrO	GPM	10%	3%	RnkSS
W_5	60	5000	RSP-CrO	GPM	10%	10%	RnkSS
W_5	60	5000	RSP-CrO	GPM	10%	15%	RnkSS
W_5	60	5000	RSP-CrO	GPM	10%	20%	RnkSS

From the test results, it can be seen that the solutions are affected by the mutation type and PMR. This investigation disclosed the parameters that result in a more stable search performance. Figure 4.39 gives the detail.

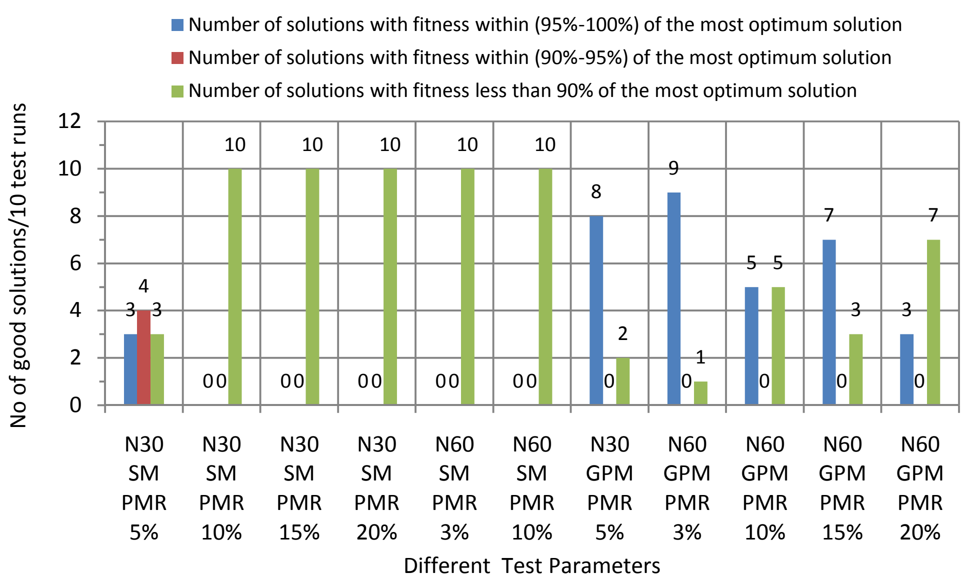


Figure 4.39: The effect of the number of SP, MM, and PMR on the search performance

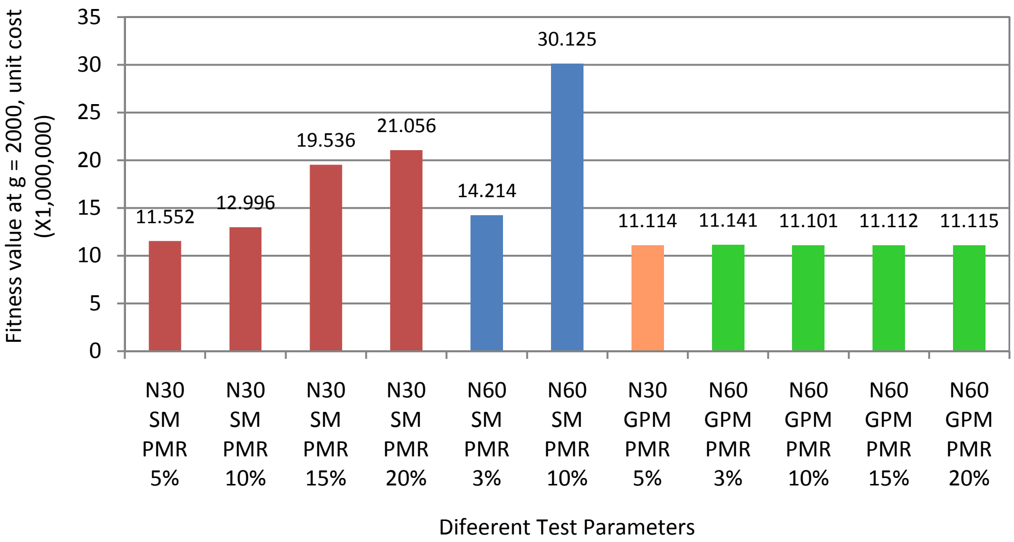


Figure 4.40: The effect of the number of SP, MM, and PMR on the best fitness value obtained

The results draw several conclusions. With SM mutation, the results show that the increase of PMR reduces the chance of obtaining good solutions (Figure 4.39) as well as decreasing the quality of the solutions (Figure 4.40). This explains that the search deteriorates with the increase of PMR. Such deterioration was expected from the applied mutation method. For instance, 20% PMR, which selects six points out of thirty station points, selects the points at random and new values are assigned at random, too. Figure 4.41 explains the idea, e.g. when a gene at location k is mutated (Figure 4.41A) (assume the new position of point k is located out of the high cost region) the genes before and after this location (k-2, k-1, k+1, k+1) need, by pure luck, to be selected and assigned new values that guarantee positions outside the high cost region at one side (Figure 4.41B). But the case that is likely to happen is somehow similar to what is shown in Figure 4.41C, where the fitness of the resulting offspring is worse than their parents and dies off before emerging into the next generation and even before the positions k-2, k-1, k+1, and k+2 are given other chances to evolve and step out of the high cost areas. As a consequence, the population is prematurely converged and stuck with no further development. Therefore, 5%-10% PMR (selecting 1 to 3 station points) at most is considered a reasonable range for SM.

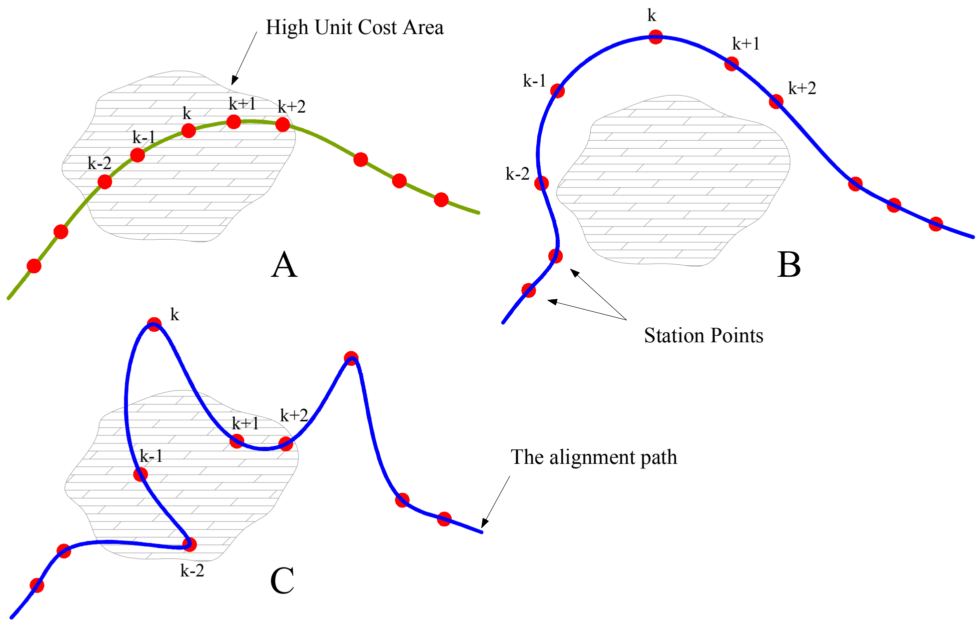


Figure 4.41: The action of point mutation rate with SM

Moreover, the result with the parameter N60 SM PMR 3% was worse than the result with parameter N30 SM PMR 5% although both select the same number of station points (1 to 2 points) for mutation. Compare the two results in Figure 4.42.

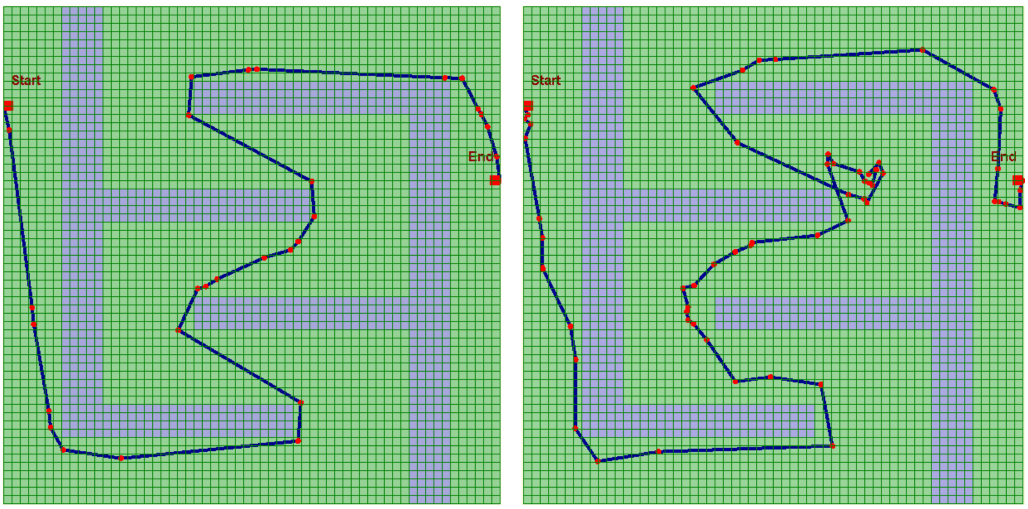


Figure 4.42: Left) a solution with N30 SM PMR 5% at G 2000, Right) a solution with N60 SM PMR 3% at G 2000

The use of GPM mutation showed different behaviour from SM. The results showed that the behaviour of the model was stable with the use of PMR up to 15%. Moreover, the mutation was proved efficient with the different number of station points (compare the results of the parameters N30 GPM PMR 5% and N60 GPM PMR 3%).

Figure 4.43 shows the best fitness of some scenarios. It shows that the convergence with N30 GPM PMR 5% and N60 GPM PMR 3% is slower than with N60 GPM PMR 10%, N60 GPM PMR 15%, and N60 GPM PMR 20%. This is due to the selection of the point ranges. For instance, N60 GPM PMR 15% selects a larger number of points at each side of the selected point to straighten the alignment compared to N60 GPM PMR 3%. This means that, in one mutation step, the alignment gets shorter with higher PMR than if less PMR is used; thus, the fitness value drops faster. For clearer understanding compare the fitness values at generation 500.

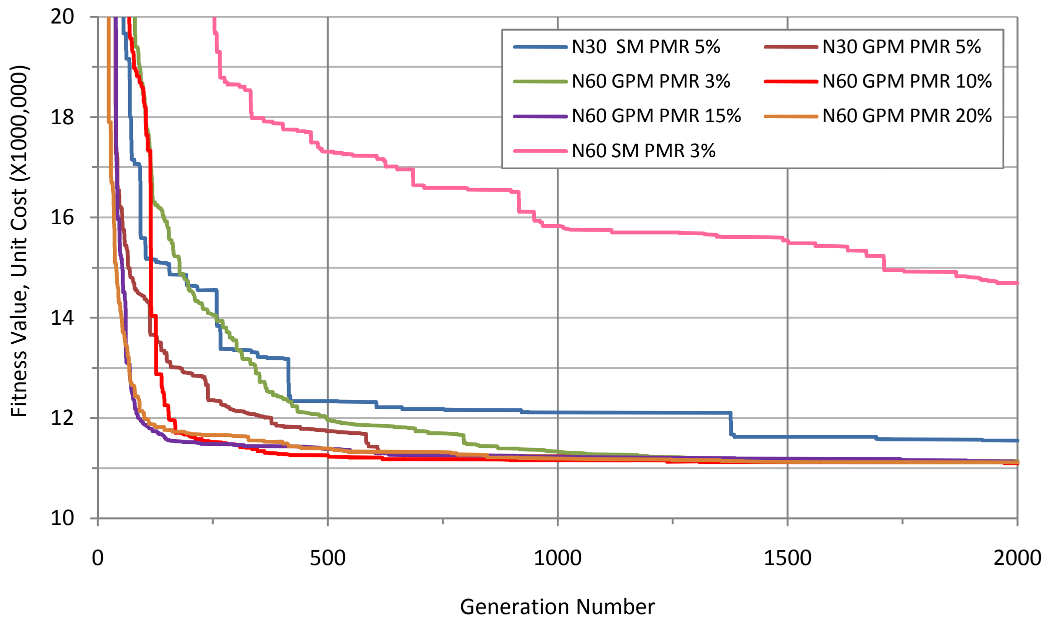


Figure 4.43: Best fitness of different test parameters

Figure 4.44 shows the most optimum alignment solutions obtained with the parameters N60 GPM PMR 10% and N30 SM PMR 10%, whereby they possess a fitness of 11,101,132 and 12,996,267 unit costs, respectively.

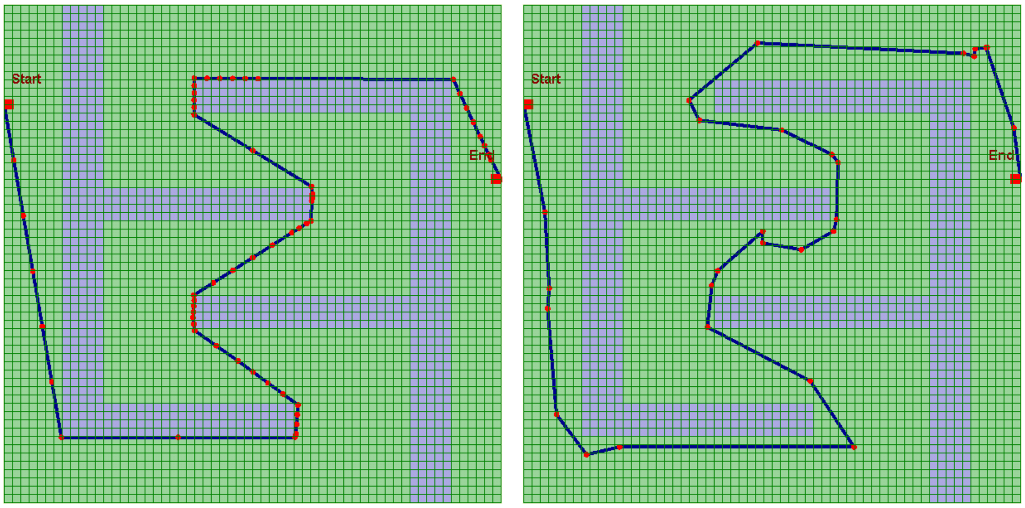


Figure 4.44: Left) The best solution with N60 GPM PMR 10%, Right) the best solution with N30 SM PMR 10%

4.5.2.4 Random Multiple Point Crossover Tests (RMPCrO)

This section is dedicated to testing the effect of several crossover methods on the behaviour of the model. The idea of multiple point crossover methods, which was inspired by the way that the solutions were represented by the use of station points, was suggested. An alignment with a relatively long length may handle different information along its length. The information is handled through the station points, and perhaps swapping information by two alignment pieces (single point crossover) may not be enough to explore many other combinations that can be formed if multiple points, as crossover methods, are used. Figure 4.45 illustrates, as an example, the mechanisms of 4-point crossover, where p_1, p_2, p_3, and p_4 are the crossover points.

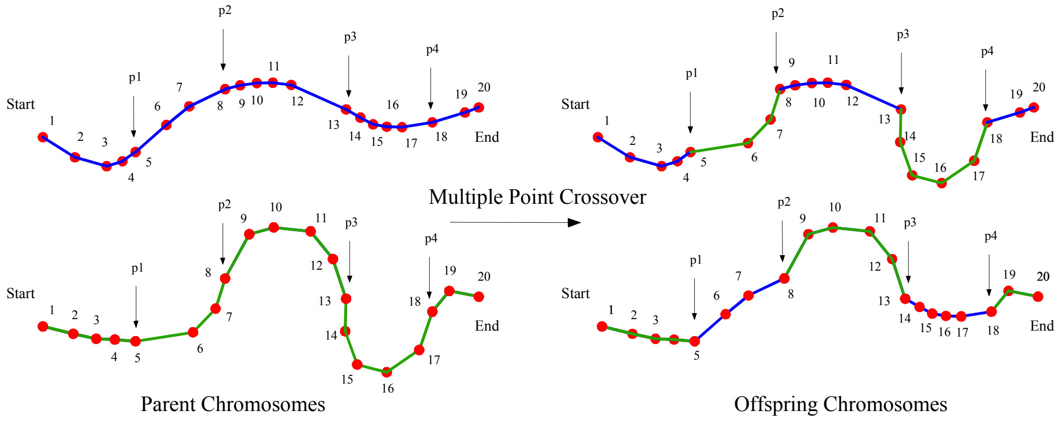


Figure 4.45: Illustration of multiple crossover mechanism

Several multiple crossover methods were made and tested in conjunction with SM upon chromosomes with lengths of 60 station points. Table 4.10 lists the test parameters with multiple-point crossover methods. These tests were set to show how efficient are the crossover methods in improving the model performance and to find the most suitable crossover method that conforms the station point representation approach.

Table 4.10: Test parameters for different crossover methods

W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS
W_5	60	5000	RSPCrO	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 2P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 4P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 6P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 8P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 10P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO – 12P	SM	10%	3%	RnkSS

Firstly, the fitness of the obtained solutions from the ten test runs was evaluated. It was found that using RMPCrO from six to ten points gives more stable results by which eight solutions out of ten test runs possessed fitness higher than 60% of the most optimum solution (the most optimum fitness value, which was obtained earlier, '11,101,132') (see Figure 4.46).

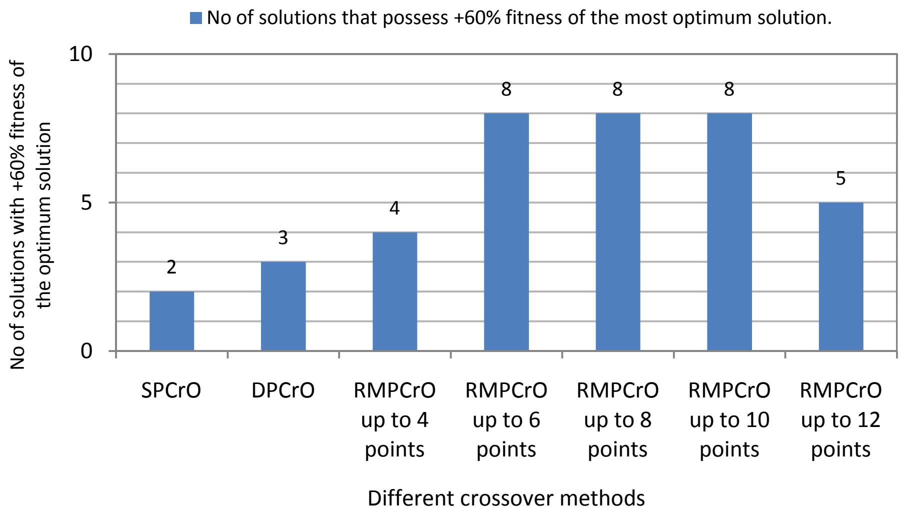


Figure 4.46: The effect of crossover methods on obtaining better solutions

Secondly, the minimum and maximum fitness were also drawn. In general, it was found that a relatively good solution was obtained from each crossover method and therefore the results cannot be judged based on the minimum fitness graph (Figure 4.47). Meanwhile, from the maximum fitness graph, as in Figure 4.48, it can be seen that the solutions that use crossover with six points and upwards are fitter solutions.

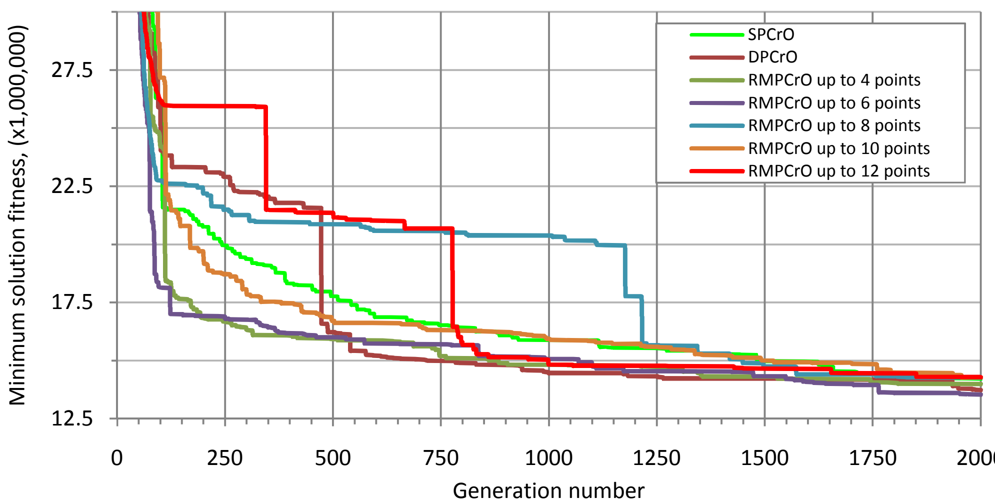


Figure 4.47: Minimum fitness graph obtained from each test scenario/10 test runs

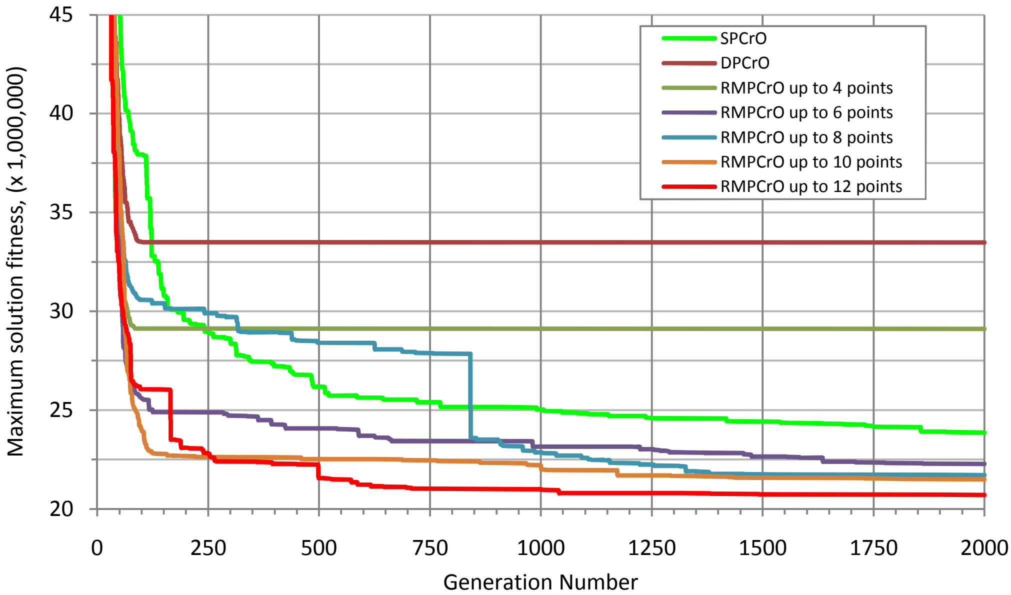


Figure 4.48: Maximum fitness graph obtained from each test scenario/10 test runs

The results reveal that random multiple-point crossover (RMPCrO) from six up to ten points performs better than the RSPCrO. Figure 4.49 shows the best results of using RMPCrO-8P (N=60) and SPCrO (N=30) at generations 2000 where they possess fitness values of 11,917,434 and 12,996,288 unit costs respectively.

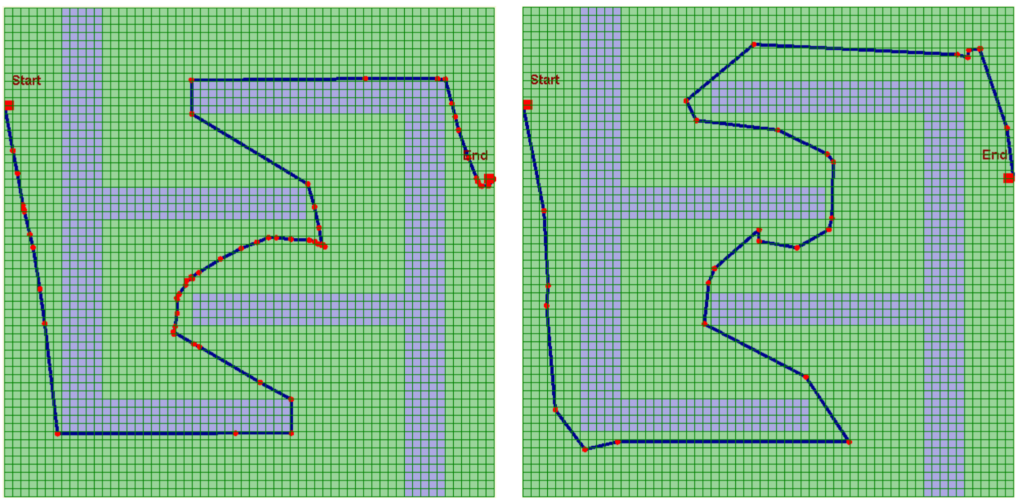


Figure 4.49: Left) the best solution with RMPCrO-8P at generation 2000 (N=60)

Right) the best solution with SPCrO at generation 2000 (N=30)

To verify the efficiency of the multiple crossover method, a test scenario was made for RMPCrO-8P with GPM. The results showed that this crossover, in addition to producing the most optimum solution with fitness 11,079,988 (Figure 4.50), also resulted in producing eight fit solutions with fitness of +98% of the most optimum result (Figure 4.51). Thus, RMPCrO-8P can be considered, as an average, as the most efficient crossover method among the tested ones.

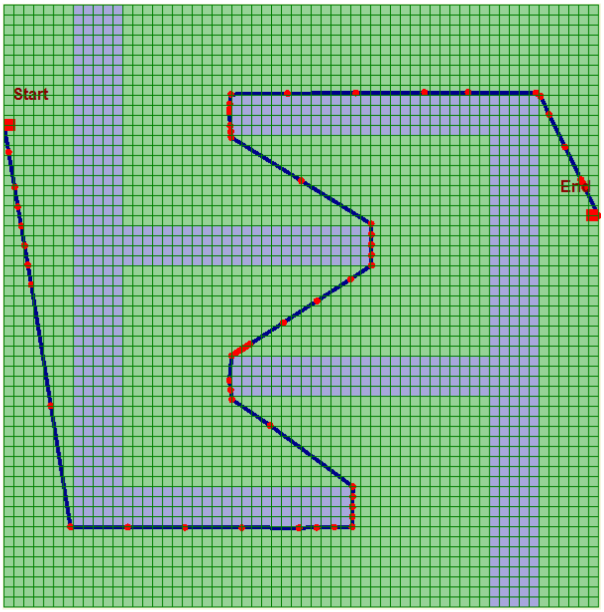


Figure 4.50: The most optimum solution using RMPCrO-8P and GPM

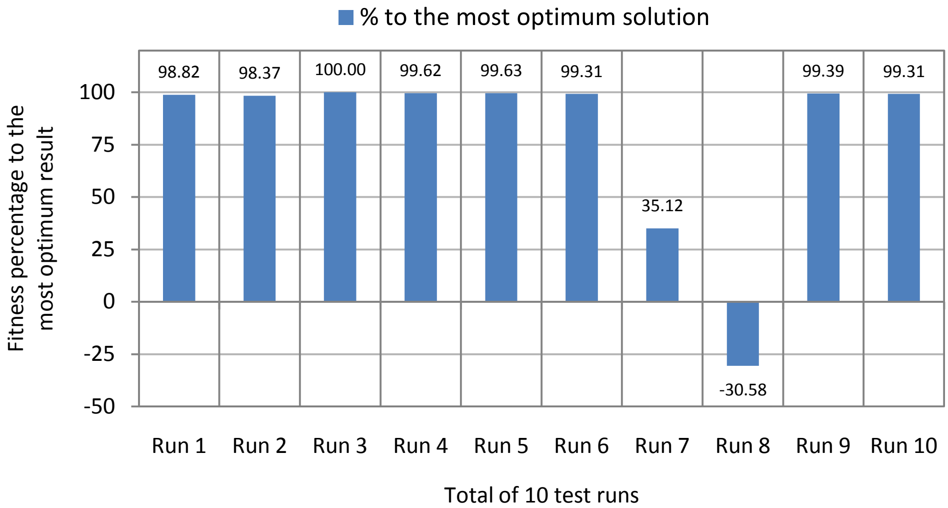


Figure 4.51: % of solution fitness to the most optimum solution

with the use of RMPCrO-8P GPM IMR 10%

4.5.2.5 Individual Mutation Probability Tests (IMR%)

Increasing the rate of good results and guaranteeing more stable performances for the model are still demanding. In this section, different rates for individual selection to undergo mutation are set. During the tests the IMR are made variable whereas the other parameters, which have been obtained from previous tests, are kept fixed. Table 4.11 shows the list of parameters, and Figure 4.52 and Figure 4.53 show the results.

Table 4.11: Test parameters of different IMR

W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS
Tests with SM							
W_5	60	5000	RMPCrO-8P	SM	5%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	SM	10%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	SM	15%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	SM	25%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	SM	50%	3%	RnkSS
Tests with GPM							
W_5	60	5000	RMPCrO-8P	GPM	5%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	GPM	10%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	GPM	15%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	GPM	25%	3%	RnkSS
W_5	60	5000	RMPCrO-8P	GPM	50%	3%	RnkSS

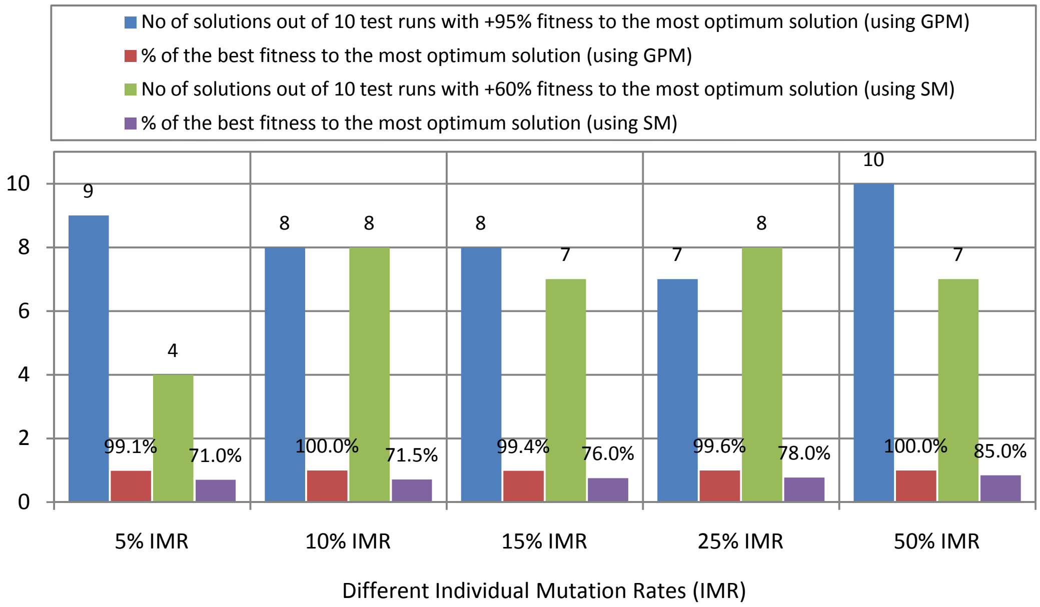


Figure 4.52: The effect of different IMR on the model performance using SM and GPM

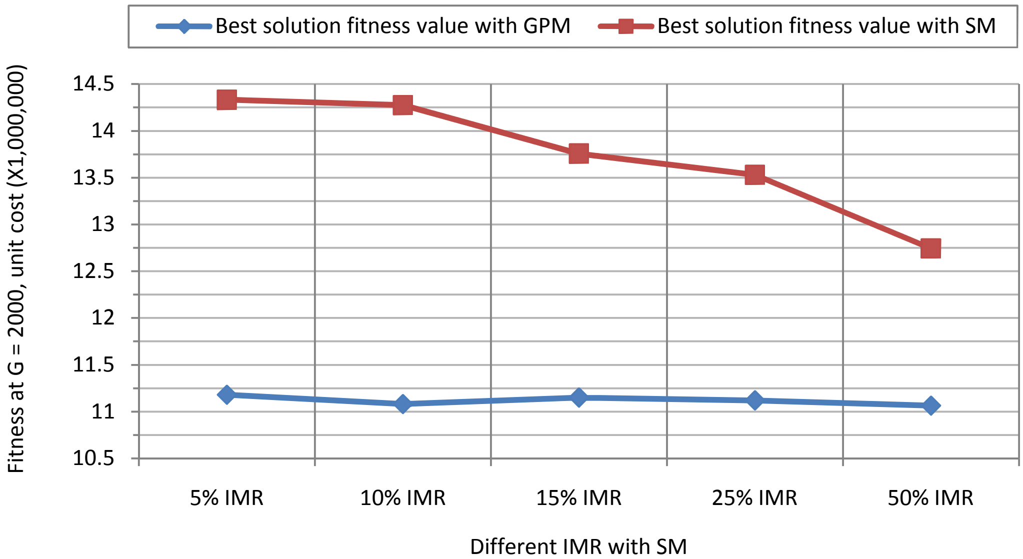


Figure 4.53: The effect of IMR on the best fitness value using GPM and SM

In general the results showed that better performances were obtained with the increase of IMR. Moreover, it can be noted that the results with SM give a clearer vision on how IMR affects the model performance. The reason for this is because SM is less efficient than GPM and therefore the effect of these parameters would be more effective and clearer with SM than with GPM.

Figure 4.54 shows the best result obtained with SM and 50% IMR.

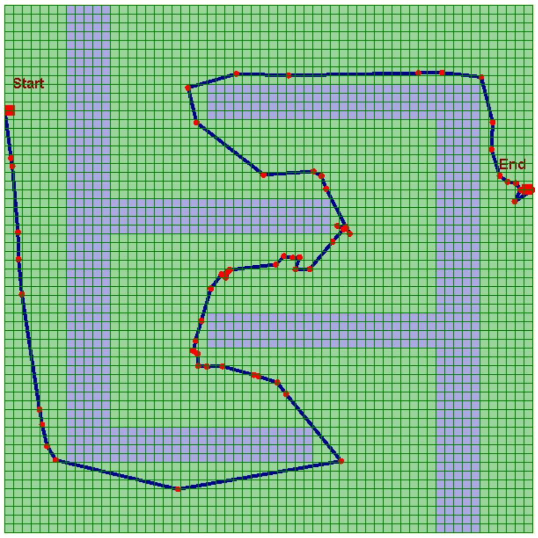


Figure 4.54: A test result at generation 2000 with SM and 50% IMR

With the above results, a question arises, why with higher IMR does the model perform better? However, this contradicts the theory of GA, which suggests low mutation probability. This may be an indication of losing the genetic diversity within the population and then maintaining the balance through mutation. This phenomenon could be critical and may put the whole process under risk. In such a case and when the mutation operator fails to make an effective search, which is likely to happen under many implicit factors, the whole process may fail to locate a good solution. The question is how further can we keep the diversity of the pool and make more efficient searches to improve the performance of the model? A bigger population size may be an answer and still be demanding for more genetic diversity in the population. However, this may overburden the process and will not explain why the solution is likely to stick at local optima. Therefore, a trend would go for understanding of why the useful information and genetic diversity are getting lost during the offspring reproduction process. The following section tries to give an answer.

4.5.2.6 Different Selection Mechanisms

The tests that have been performed so far were based on the Ranking Selection Scheme (RnkSS) by which the individuals are selected depending on their ranks in the population (see Section 4.3.4.3). Using RnkSS, the first best individual (rank 1) and the second best (rank 2) are selected for reproduction and the same for the third and fourth ranked individuals. This approach is likely to give higher chances to the good individuals at the top of the population to produce better offspring than those ranked at the bottom of the population. This explanation tells that the contribution of the solutions with high fitness values to produce good offspring could be low. However, the high fitness valued individuals might hold, in some parts of their chromosomes, important and useful information about the optimum solution. Moreover, mating those individuals ranked one after each other may also produce individuals that are radically similar to their parents. Contextually, individuals with similar fitness (although it is not always the case) may share similar characteristics and genetic information which, after a short time from the start of the process, would probably create identical offspring. Having this phenomenon, the entire population would quickly be dominated by the offspring resulting from the top parents and the process ends up with a premature convergence leaving the chance only for mutation to maintain the diversity and bring the lost information back into the pool through making new searches.

To give some practical explanation to what has been mentioned, the most efficient test parameters that have been found so far were used to test the effect of different selection schemes on the model performance, population convergence, and obtaining optimum solutions.

In this section, the reverse selection scheme (RevSS; see Section 4.3.4.3 for detail) was tested and compared with RnkSS. This selection scheme selects the best parent (rank 1) to breed with the last one (rank n), where n is the population size, and the second rank to breed with rank n-1, and so on. Table 4.12 summarises the test parameters for RevSS investigation.

Table 4.12: Test parameters associated with RevSS

W	N	P_{size}	CrO_Type	MM	IMR%	PMR%	SS
W_5	60	5000	RMPCrO – 8P	GPM	10%	8%	RevSS
W_5	60	5000	RMPCrO – 8P	SM	10%	3%	RevSS

The results of the fitness percentage to the most optimum solution (fitness: 11,079,988 as the most optimum fitness that has been obtained from the tests in section 4.5.2.4) of each test run are shown in Figure 4.55. It can be seen that the performance of the search is improved and significantly affected by RevSS.

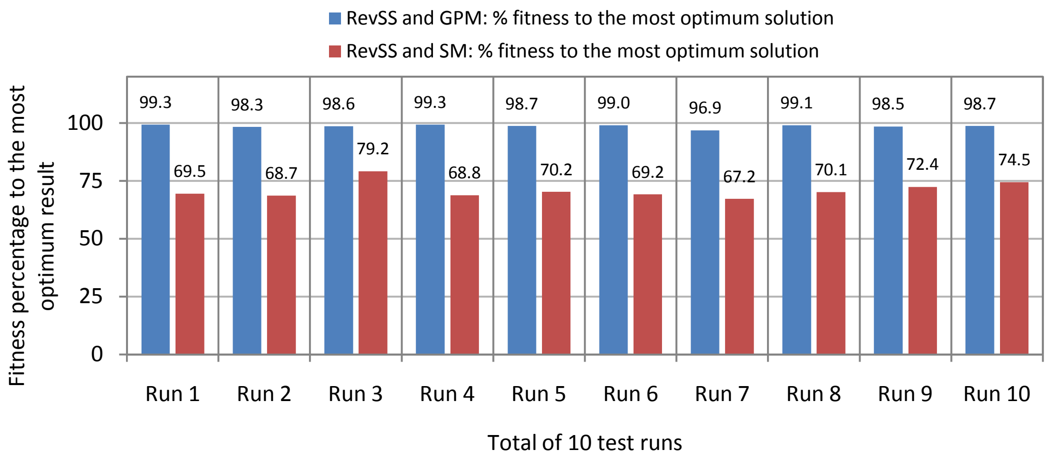


Figure 4.55: The effect of selection schemes on the stability of the search performance

The results show that RevSS with GPM and RevSS with SM resulted in producing 10 solutions out of 10 test runs with fitness higher than 97% and 65% of the most optimum solution, respectively. These were considered better performances compared with the results when RnkSS was used. The latter produced 8 solutions out of 10 test runs with fitness higher than 97% for GPM and 60% for SM of the most optimum solution. This indicates that RevSS outperforms RnkSS due to the explanations given earlier. This gives great support for those who suggested incest prevention to avoid premature convergence and getting stuck at local optima (Eshelman & Schaffer, 1991).

Figure 4.56 and Figure 4.57 show the convergence behaviour of GPM and SM with the two selection schemes for two of the best solutions from each method. It can be seen that RevSS has a slower convergence due to keeping the diversity of the genetic information for longer times. Furthermore, the low variations among the different test runs prove the robustness of the parameters to search for a good solution.

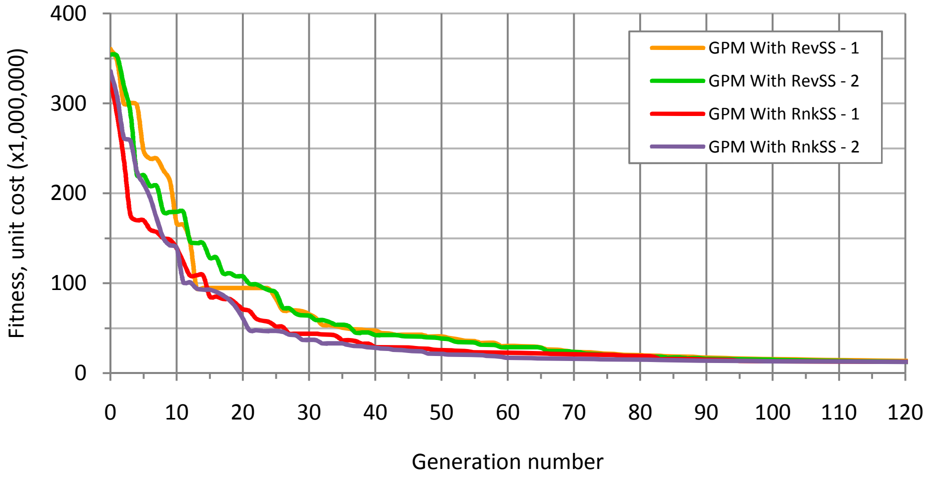


Figure 4.56: Convergence comparisons for RevSS and RnkSS with GPM

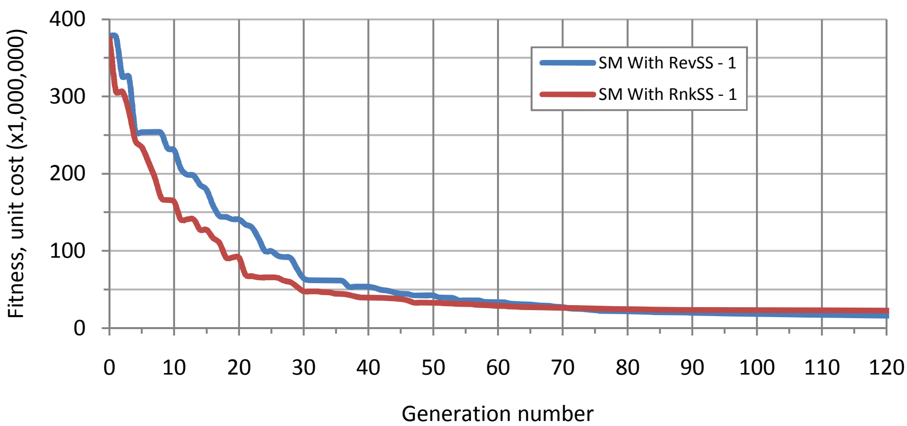


Figure 4.57: Convergence comparisons for RevSS and RnkSS with SM

The results proved that the performance of the model could perhaps be more stable with the use of RevSS than using RnkSS.

Figure 4.58 shows an optimum solution obtained by using GPM, 10% IMR, 8% PMR, and RevSS.

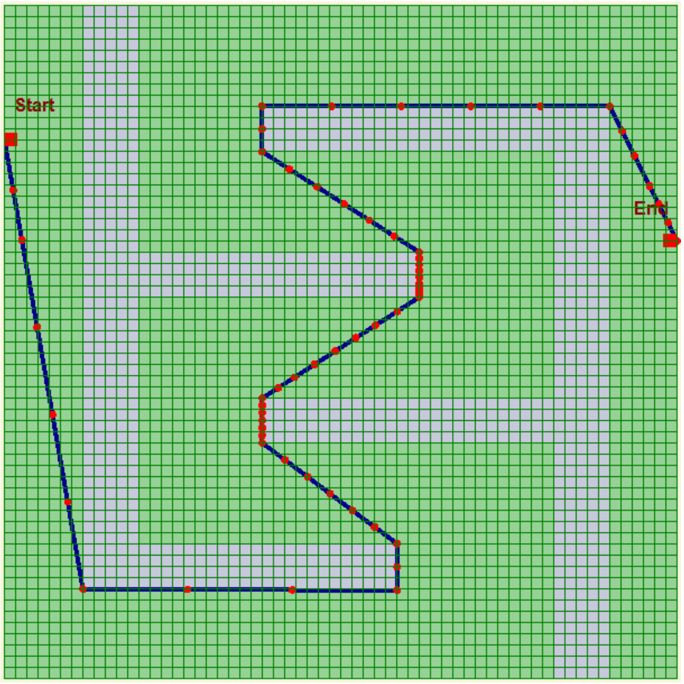


Figure 4.58: An optimum solution obtained with 10% IMR 8% PMR GPM and RevSS

It should be mentioned that the above tests have complemented and revealed the shape of the relations among the different tested parameters.

Table 4.13 below summarises the parameter values that have been found so far, which hold promises for an efficient search model to find a global or near global 2D highway alignment solution.

Table 4.13: Summary of the tuned parameters for a 'good' 2D alignment solution

P_{size}	N	CrO_Type	MM	IMP	PMR	SS	Possible good results
5000+	60	RMPCrO-8P	GPM	5%-10%	3%-15%	RevSS	9±1/10 within +98% of the most optimum solution
5000+	60	RMPCrO-8P	SM	5%-10%	3%	RevSS	9±1/10 within +60% of the most optimum solution

4.5.3 TEST GROUP 3: TESTS ON RANDOMLY GENERATED WORLDS

The parameters that have been found so far are based on specific world cases. These parameters have been tuned using plain worlds (W_1) and (W_3), plain worlds with a high cost feature (W_2), a world with zigzag features (W_4), and a world with backbend requirements (W_5). The search capability of the model was tested and tuned on these worlds considering they are representing critical cases that might occur in real world applications.

4.5.3.1 Random Worlds

The reliability of the model parameters in real world applications needs further verifications. For this purpose an algorithm was developed to generate random worlds. A random world can be defined as a world that can constitute N number of different land features that are randomly distributed within the study area. Different land features of a world may generate different location and construction costs due to the land use characteristics (e.g.: environmentally sensitive areas, political areas, archaeological areas etc) and soil type properties.

Random worlds may form cases that are more difficult than specific worlds and even from real world problems due to the inconsistency in distributing the land features. However, the aim is to find parameter values that suit different cases (specific and random worlds) that guarantee finding an optimum or near optimum solution when they are applied in real world problems. Figure 4.59 is an example of a randomly generated world (RW_6).

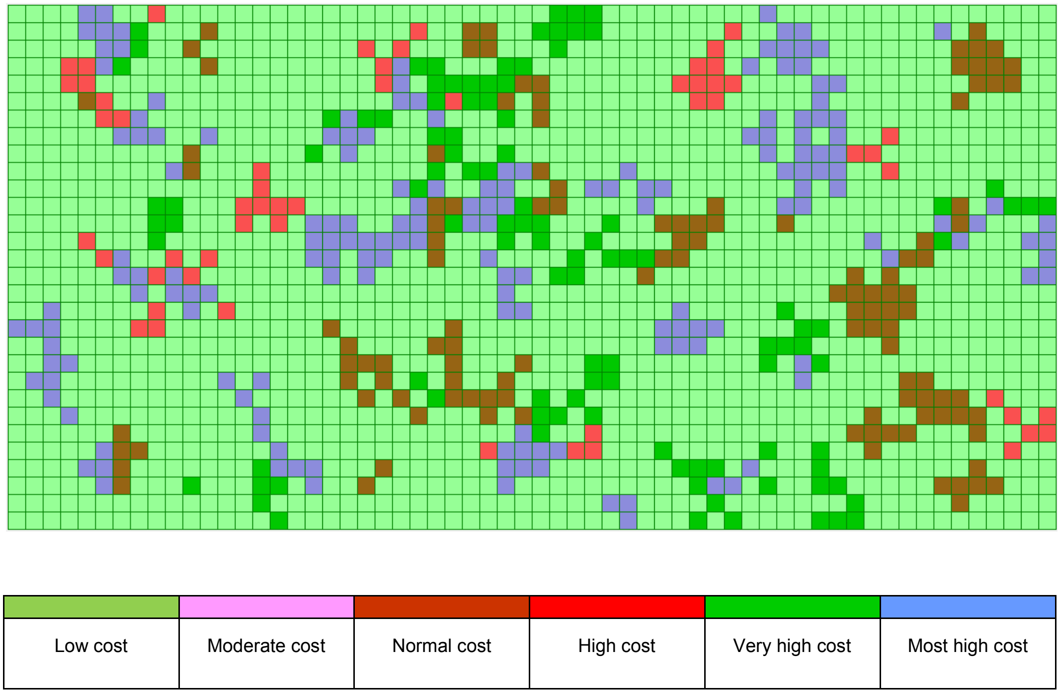


Figure 4.59: A randomly generated world (RW_6) with different land features

This world is 12000m x 6000m of a grid cell size 200m x 200m (1800 cell number). The different colours represent different land features. Each land feature is assigned an average unit cost varying between low costs of 100 unit/m^2 (the light-green coloured cells), and high costs of 7000 unit/m^2 (the blue coloured cells). These unit costs may represent different cost configurations. For instance, a combined location cost of a land may be represented as a sum of land acquisition cost and treatment costs when the soil has poor condition, and/or when the location has an environmental value like green areas or marshes.

On this random world, several parameters were tested and all the parameters were taken from the good results obtained earlier. The test parameters are as in

Table 4.14 and the rate of the good results that are optimum or near optimum are shown in Figure 4.60.

Table 4.14: Random world (RW_6) test parameters

P_{size}	N	CrO_Type	MM	IMP	**PMR**	SS
5000+	60	RMPCrO-8P	GPM	8%	**8%**	RevSS
5000+	60	RMPCrO-8P	GPM	8%	**8%**	RnkSS
5000+	60	RMPCrO-8P	GPM	8%	**12%**	RevSS
5000+	60	RMPCrO-8P	GPM	8%	**12%**	RnkSS
5000+	60	RMPCrO-8P	GPM	8%	**15%**	RevSS

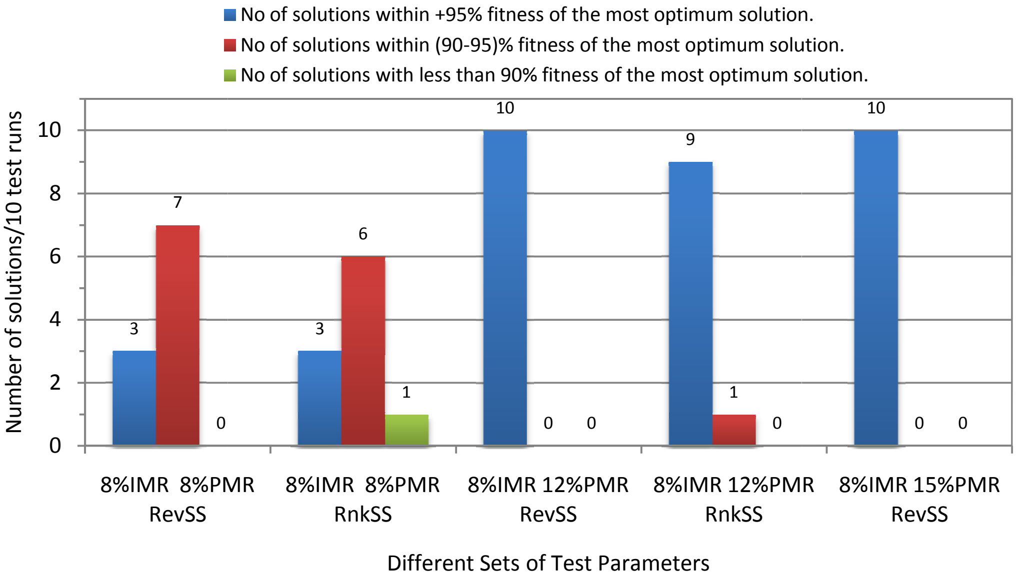


Figure 4.60: Number of good solutions of different test parameters

To explain how well the different parameters were behaving, minimum and maximum fitness from the ten test runs per parameter are drawn in Figure 4.61 and Figure 4.62. The lower the variation in the fitness value the more stable is the parameter provided that the parameter has a high rate of good results (refer to Figure 4.60). It can be seen that the least variation is obtained when RevSS is used with 12% PMR and 15% PMR. This explains that the model behaviour is more stable when higher PMR is used compared with the world W_5. This may happen due to the configuration of the land features where wider areas of high costs cluster at one location. The alignment at such locations requires wider jumps and this is accomplished by using higher PMR.

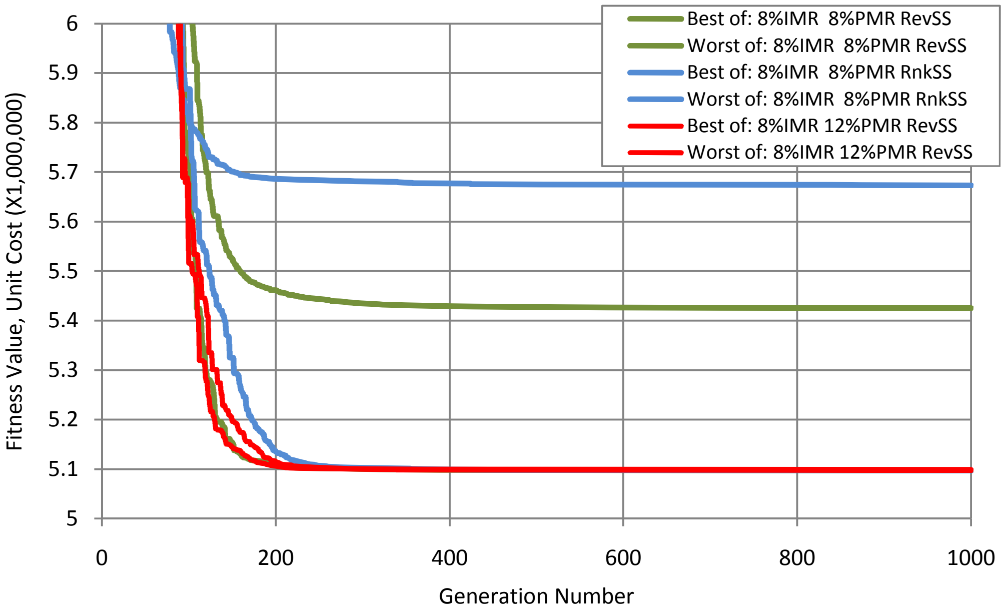


Figure 4.61: Minimum and maximum solution fitness with different parameters

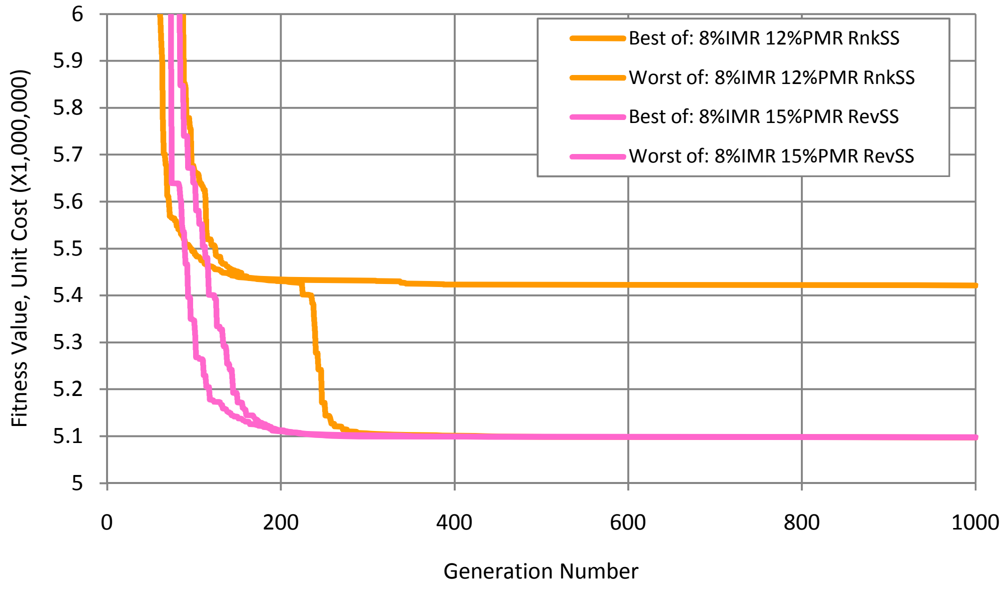


Figure 4.62: Minimum and maximum solution fitness with different parameters

Comparing the variations between 8% IMR 8% PMR RevSS and 8% IMR 8% PMR RnkSS, it can be seen that incest prevention is still valid (RnkSS has produced higher variation). From Figure 4.60 and the variation graph, it can be seen that the parameter sets of 8% IMR 12% PMR RevSS and 8% IMR 15% PMR RevSS are the most efficient sets to produce least variations and 10/10 solution within +95% fitness to the most optimum value. Figure 4.63 shows the most four-variance test results from the ten test runs obtained from the parameter set 8% IMR 12% PMR RevSS, where they all come up with almost the same solution.

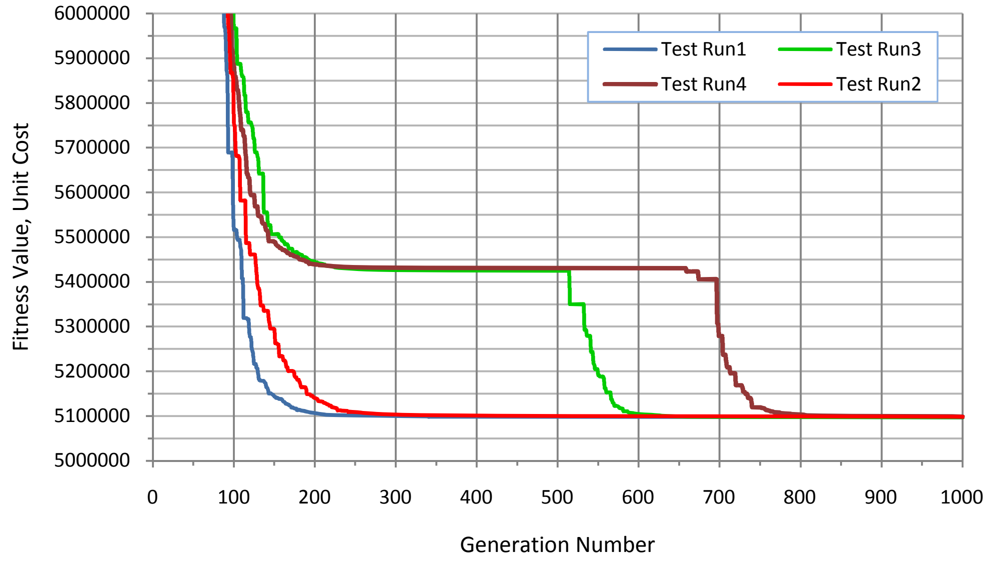


Figure 4.63: Fitness graph for different test runs of the same parameter set (8% IMR 12% PMR RevSS GPM MPCrO-8)

Figure 4.64 shows an optimum solution obtained from parameter set 8% IMR 12% PMR RevSS. The solution has the following characteristics:

- as short as possible
- skirts the high cost areas
- as straight as possible at the locations where no bend is required
- relatively smooth.

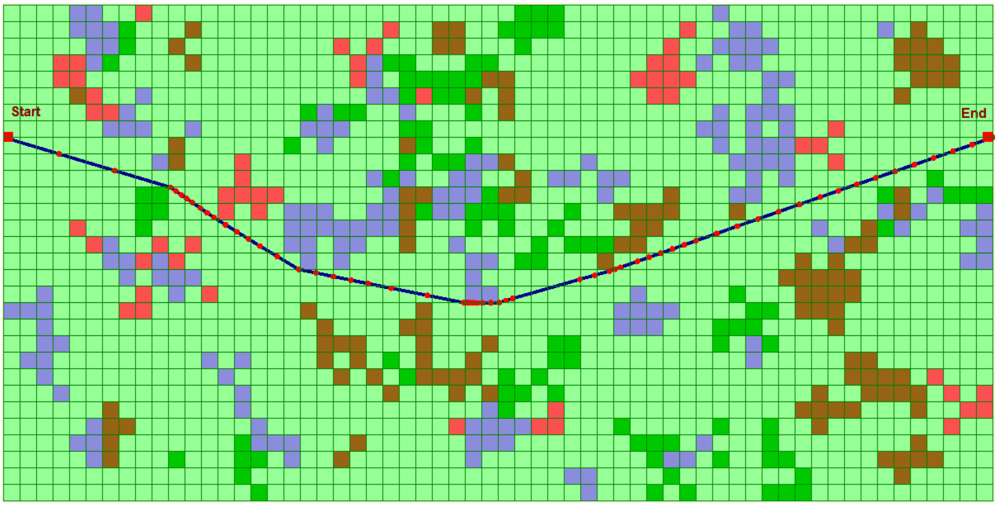


Figure 4.64: An optimum solution of a parameter set (8% IMR 12% PMR RevSS GPM MPCrO-8)

Table 4.15 shows a simple statistical analysis for the results of 10 test runs for the parameter set 8% IMR 12% PMR RevSS GPM MPCrO-8P.

Table 4.15: A statistical analysis for the best solution set

Test run	Fitness value (Unit Cost)	% Variance from optimum value
1	5096976	0
2	5097525	0.011
3	5097265	0.006
4	5098862	0.037
5	5097192	0.004
6	5097625	0.013
7	5098132	0.023
8	5097133	0.003
9	5098329	0.027
10	5097190	0.004
Average	5097623	0.0127
STD	620	0.0122

4.5.3.2 Tests on Specific-Random Worlds (SRW_7)

The random world RW_6 is manually edited to create a new world so that it hosts both specific and random world features. This new world, which is denoted as SRW_7, is used to test and verify the most meaningful parameters that have been found. The aim with this world is to combine zigzag features and the requirement for a backward alignment within different land features that are randomly distributed in the area under consideration.

The intension of testing and tuning the model parameter values on different worlds is to ensure that the model is tested and applicable on as many world cases. Figure 4.65 is the world example SRW_7 that combines random and specific world features. Table 4.17 shows the parameters that are used to test the model on the new world. With the listed test scenarios, the effects of RSA and RDA as two strategies with the applied mutation (see Section 4.3.4.5 ii) are also compared.

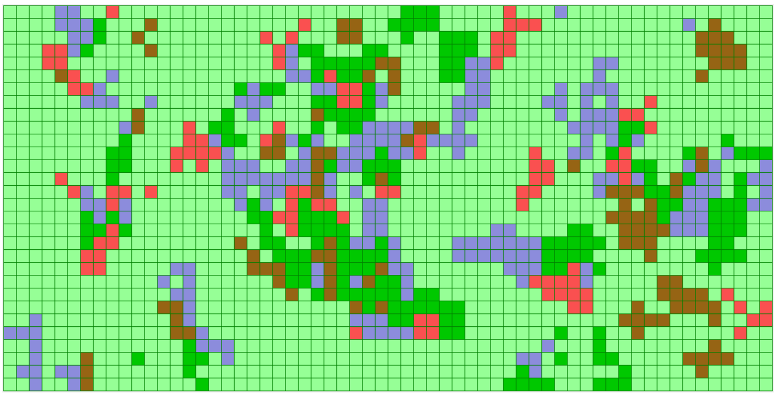


Figure 4.65: A manipulated random world (SRW_7)

Table 4.16: Different test parameter values used with SRW-7

Scenario	P_{size}	CrO_Type	MM	IMR	PMR	SS	G
S1	5000+	RMPCrO-8P	GPM (RSA) *	8%	8%	RevSS	1000
S2	5000+	RMPCrO-8P	GPM (RDA) *	8%	8%	RevSS	1000
S3	5000+	RMPCrO-8P	GPM (RDA)	8%	10%	RevSS	1000
S4	5000+	RMPCrO-8P	GPM (RDA)	8%	12%	RevSS	1000
S5	5000+	RMPCrO-8P	GPM (RDA)	8%	10%	RnkSS	1000

(*) refer to Section 4.3.4.5 (ii) for details on RSA and RDA.

Figure 4.66 presents how stable are the parameters in producing optimum or near optimum solutions from the ten test runs per scenario.

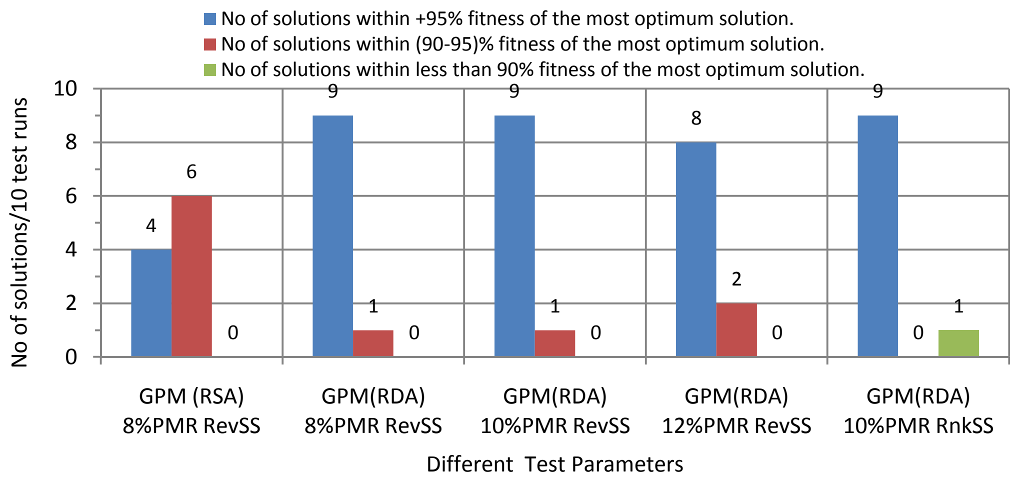


Figure 4.66: The effect of different parameters on the performance of the model

From investigations based on the results it can be concluded that using RDA as a strategy with GPM outperforms RSA strategy. Using GPM with RSA (S1) resulted in producing four solutions within +95% fitness of the most optimum fitness value while RDA guaranteed 9/10.

Table 4.17 shows the average and standard deviation (STD) of the fitness values and compares the different scenarios with each other. Figure 4.67 shows the convergence of the best solutions from each set.

Table 4.17: Average and STD of the test results

	S1	S2	S3	S4	S5
Run 1	6,855,432	6,809,786	6,801,728	6,852,179	6,685,719
Run 2	7,056,026	6,783,503	6,828,801	6,831,110	6,787,102
Run 3	6,921,657	6,831,183	6,821,040	6,662,772	6,671,974
Run 4	7,029,012	6,795,050	6,831,609	7,064,719	6,816,653
Run 5	6,852,086	6,806,760	6,793,140	6,636,645	6,625,287
Run 6	7,096,529	6,819,823	7,055,033	6,843,865	6,799,688
Run 7	7,066,831	6,902,739	6,804,593	6,809,266	6,798,703
Run 8	7,094,643	6,799,396	6,818,268	6,627,839	6,799,688
Run 9	7,075,065	6,798,163	6,793,187	6,804,159	7,536,917
Run 10	6,926,126	6,971,461	6,789,132	6,974,090	6,793,304
Average FF	**6,997,341**	**6,831,786**	**6,833,653**	**6,810,664**	**6,831,504**
STD	**98,127**	**59,309**	**79,272**	**141,468**	**256,815**

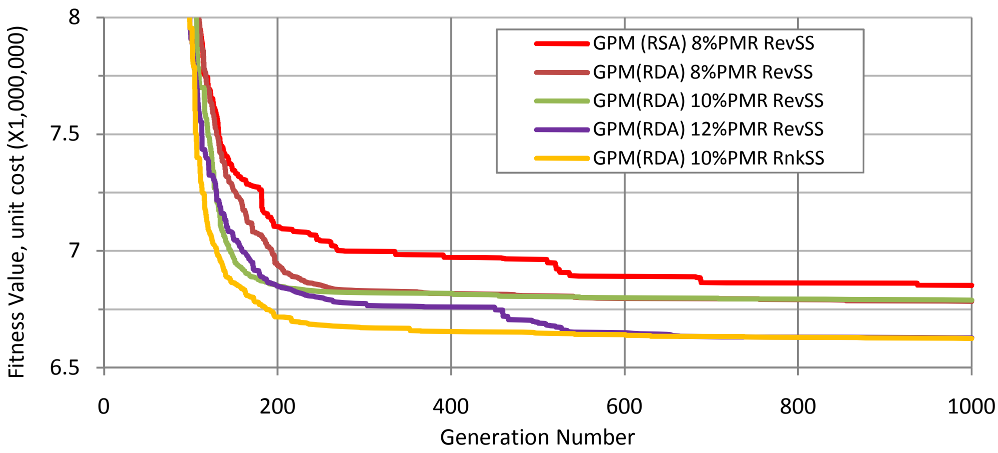


Figure 4.67: Best fitness result from different parameters

From the above result presentations it can be seen that the performance differences between the different parameters are small. Therefore, all the parameter values can be accepted within the planner and decision makers' viewpoints and project requirements.

Figure 4.68 shows an optimum solution result obtained from S3. This solution possesses a fitness value 6,789,132 unit cost which is as 97.52% optimum as the most optimum solution. Figure 4.69 shows a result (which is not the optimum) obtained from test S1 in which it handles a fitness value equal to 7,029,011. These two presentations show how the model is efficient to find fitter solutions within narrower corridors and through irregular land features.

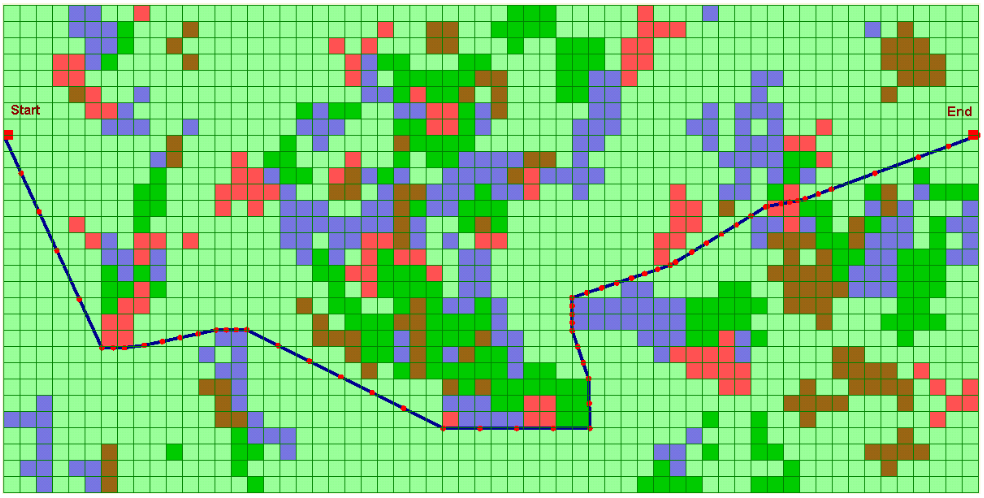


Figure 4.68: An optimum solution from test scenario S3 at generation 1000

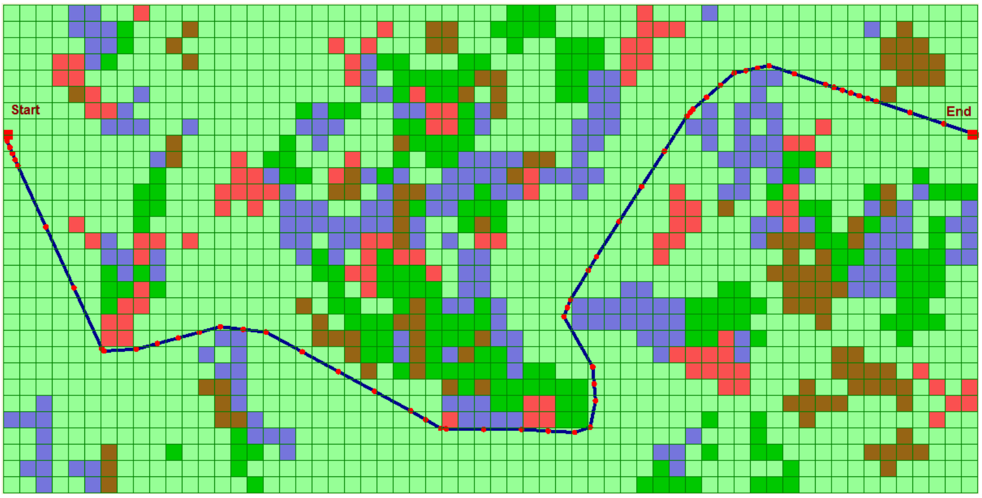


Figure 4.69: A solution from test scenario S1 at generation 1000

The tests on different world categories proved that the parameter limit values found earlier (Table 4.13) are still valid and effective to find solutions that are optimum or near optimum. However, using RDA as a strategy with the applied mutations makes the parameters more valid. Moreover, when the alignment does not require possessing a backward bend, the more efficient parameter for PMR may vary between 8%-15% instead of 3%-15%. Table 4.18 summarises the results.

Table 4.18: Result summary

P_{size}	N	CrO_Type	MM	%IMP	%PMR	SS	Alignment feature	No of good solution[1]
5000+	60	RMPCrO-8P	GPM (RDA)	5%-10%	3%-15%	RevSS	With backward bends	9 ±1/10
5000+	60	RMPCrO-8P	GPM (RDA)	5%-10%	8%-15%	RevSS	More straight-forward	9 ±1/10

[1] Note: ±1 is left for the randomisation characteristics of GA process.

4.6 SUMMARY

The aim behind the formulations presented in this chapter was to test the applicability of the proposed design approach of station points for horizontal highway alignment development (2D) and the possibility of using GA to solve the optimisation of the problem. For these purposes a GA-based model was formulated so that an optimum horizontal alignment is searched through finding optimum locations of the station points in a horizontal 2D plane (XY plane). It is also worth mentioning that the study had to set and tune the model parameters required to run the optimisation process for optimum or near optimum solutions.

The model was tested using several world scenarios. The worlds were in different sizes and they varied between worlds with specific land features with absolute known solutions and worlds with random land configurations. Developing worlds with known solutions were important to evaluate the techniques employed to produce optimum solutions. It should be mentioned that the accuracy of the solutions obtained was constrained by the grid cell sizes and the number of station points representing the study area and the alignment, respectively. Furthermore, it was assumed that the alignment consists of a number of straight horizontal segments connecting the subsequent station points sequentially. It should also be mentioned that the study area was considered rectangle with regular square grid cell sizes (with user-defined dimensions) and it was assumed that the characteristics of the locations represented by the grid cells were unique (uniform) within the boundary of each grid cell area.

The model used a fitness function based on the weighted addition of the factors which the users considered important. The evaluation took the location and construction dependent costs into consideration so that the solution avoids high cost fields and reduces the length for less location and construction costs, respectively.

The model developed in this chapter is only able to produce horizontal highway alignments with piecewise linear trajectory characteristics possessing relatively sharp horizontal bends. This is because no criteria, as constraints, were embedded to define geometric characteristics for smoother and more realistic horizontal alignment consideration.

Based on the solution results and the characteristics of the solutions it can be concluded that the model, using the notion of station points, was successful in producing optimum horizontal piecewise linear trajectory alignments (rough 2D alignment), and the search technique (the GA) was able to make efficient searches and find the global optimum or near optimum solution.

However, as the highway alignment is a 3D problem, a highway alignment solution cannot therefore be optimum unless the third dimension, which is represented by elevations or Z coordinates, is simultaneously considered with the 2D coordinates of the alignment. Therefore, the final verification of the test parameters needs further formulations, tests, and verification to find the most effective set that determines an optimum 3D highway alignment. The next chapter is all about 3D highway alignment formulation.

CHAPTER 5

AN OPTIMISATION MODEL FOR THREE-DIMENSIONAL (3D) HIGHWAY ALIGNMENTS

5.1 INTRODUCTION

Practically, three-dimensional (3D) highway alignment is represented by horizontal and vertical alignments. Horizontal alignment is the projection of the 3D alignment in the XY plane, while vertical alignment represents the heights or elevations along the centreline of the horizontal alignment in the XZ plane and it is also known as the alignment profile.

In Chapter 4, a GA-based model for 2D highway alignment development was successfully built and, therefore, this part of the study is dedicated to investigate the possibility of extending the model to deal with 3D highway alignment using the station point approach.

The concept of station point for highway alignment configuration in a 3D space is investigated in this chapter. Using the proposed design approach of station points, both alignments are configured by X, Y, and Z coordinates of the station points.

This chapter starts with defining the new approach for developing a 3D highway alignment. The required study format to build a 3D model is then described. Thereafter, the costs associated with highway alignment profile are detailed and integrated with the model. Moreover, possible formulations for the GA operators are discussed. The chapter also demonstrates different test scenarios and culminates with the results and main conclusions.

5.2 THE NEW DESIGN APPROACH FOR 3D HIGHWAY ALIGNMENT

The same definition idea of station point approach for 2D highway alignment was extended to include the third dimension for 3D highway alignment representation. The third dimension of an alignment was configured by assigning heights (Z coordinates) to the X and Y coordinates of the station points.

To represent a vertical alignment, the X and Y coordinates of the 2D highway alignment's station points were used to determine the horizontal distances (d) between the successive station points along the centreline of the horizontal alignments. A horizontal distance d_i between any two points was calculated according to the following equation:

$$d_i = \sqrt{(X_{i+1} - X_i)^2 + (Y_{i+1} - Y_i)^2} \tag{5.1}$$

Then, a coordinate system for the stations (horizontal distances) versus elevations was needed so that the natural ground and road elevations along the horizontal alignment centreline could be represented. The X axis represented the station distances (d), while the Z represented elevations where, at each distance (di), the height of the corresponding station point was drawn and sequentially interconnected with the other station points by which the profile of the alignment was obtained.

The horizontal axis can begin with the start station of the alignment (commonly starts at 0.00 station), and the origin of the elevation axis (Z coordinate axis) is either determined from the minimum possible elevation or it is predefined by the planner as a datum line. The maximum range for the Z axis is either determined from the maximum possible elevation or predefined by the user.

When the origin is defined as X_{start}, Z_{min}, then the position of any station point along the vertical alignment is calculated by their cumulative horizontal distances (station distances) from X_{start} and the elevation of that point. For instance, the position of point (*i*) was determined as:

$$S_i = \sum_{p=0}^{p=i-1} d_p \qquad (5.2)$$

Where S_i is the cumulative horizontal distance from the start point to point i. Then the coordinate of station point i would be S_i, Z_i.

Figure 5.1 illustrates how the station points were configured in the 2D plane and how the station distances (d) between the successive station points along the centreline of the alignment were represented. Figure 5.2 explains how the station distances, by adding heights to the station points, were converted to alignment profile.

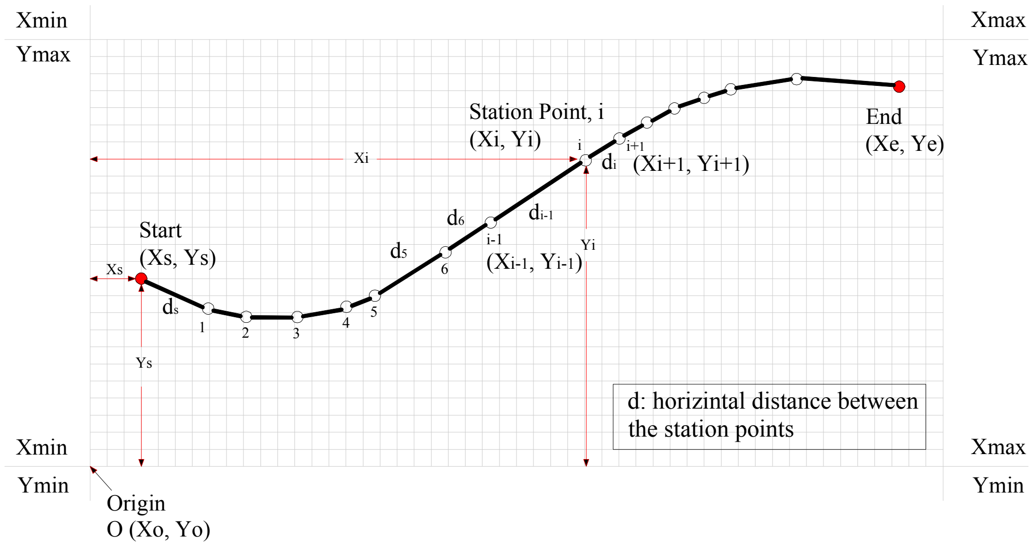


Figure 5.1: Horizontal distance representation along the centreline of a 2D highway alignment

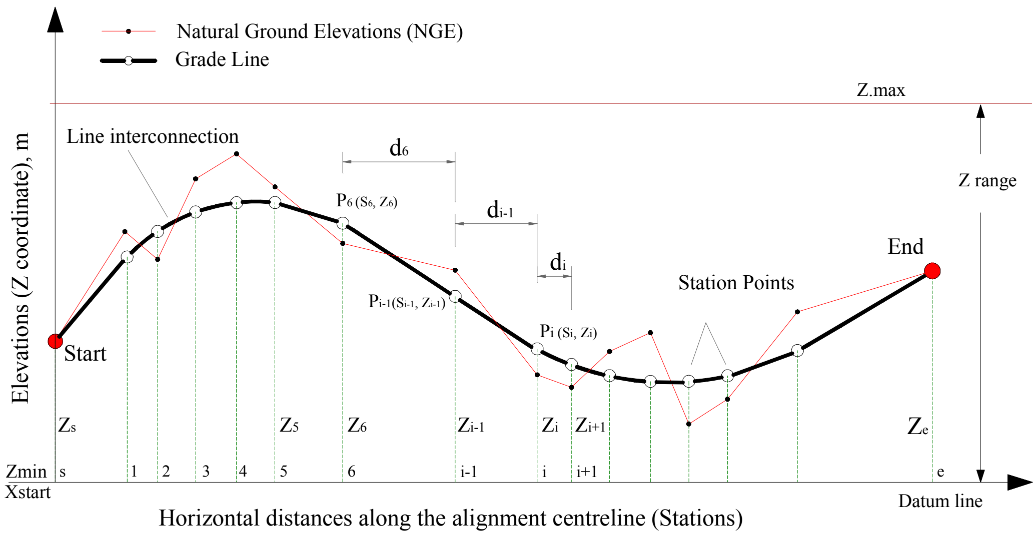


Figure 5.2: Vertical alignment representation using station point approach

Vertical alignment, similar to horizontal alignment, may require more points at changes where an upward grade is followed by a downward grade or vice versa. The demand for more points at these locations is to produce a smoother transition between the different grade line intersections.

The existing traditional method of highway alignment planning and design generates alignment profile based on a pre-generated horizontal alignment which rarely makes the final solution optimum. Moreover, the traditional approach uses the geometric design elements like VIP, tangents, curves, grades, etc. to generate the profile. The proposed new approach tries to generate a 3D highway alignment directly through the X, Y, and Z coordinates of the station points and simulteneously produce optimum horizontal and vertical alignments. This approach does not make full use of the traditional design elements; alternatively, a global optimum solution is sought through finding optimum positions for the station points.

5.3 THE MODEL FORMULATION

5.3.1 THE STUDY BOUNDARY

The same study format of the 2D highway alignment was re-formatted in such a way that the topography of the area under consideration was also included. An average ground elevation value was assigned to each grid cell of the study area. The average grid cell or zone elevation value concept was first adopted by Roberts (1957) as 'smoothed elevations' and also used by Parker (1977), Jong (1998), and Kang (2008) for cut and/or fill quantity estimations. A two dimensional matrix was developed to handle the ground elevation values. The dimensions of the elevation matrix were designed equivalent to the dimensions of the study area defined by the number of horizontal and vertical grid lines. As for the 2D alignment, the dimensions of the elevation matrix were functioned by the width, height, and grid cell size, so that:

$$k = \frac{W}{b} \quad , \qquad n = \frac{H}{h} \tag{5.3}$$

Where k is the number of vertical grids (columns), n is the number of horizontal grids (rows), W is the width of the area, H is the height of the area, and b and h are the grid cell dimensions.

Then the dimensions of the elevation matrix $M_{Elevation}$ would be (k,n) as below:

$$\mathrm{M}_{\mathrm{Elevation}}\,\{\mathrm{k},\mathrm{n}\} = \begin{bmatrix} 1 & 2 & 3 & \dots & \mathrm{k} \\ 2 & \dots & & \dots & \dots \\ 3 & \dots & & \dots & \dots \\ \dots & \dots & & & \\ \dots & \dots & & & \\ \mathrm{n} & \dots & & \dots & \mathrm{k},\mathrm{n} \end{bmatrix} \tag{5.4}$$

Such data models could be retrieved from a GIS database of the area under consideration. Each element value of the matrix represents the average elevation of the ground of the corresponding grid cell. The values of the elevation matrix were used to calculate the amount of earthwork for both cut and fill locations where the alignment traverses. In this study the topography of the study area was represented by either digital representation or contour map, as shown in Figure 5.3. A surface terrain elevation can also be plotted to make the topography of the area more imaginable, as in Figure 5.4. The surface terrain format was used only to give a prospective view and was not used for solution presentation.

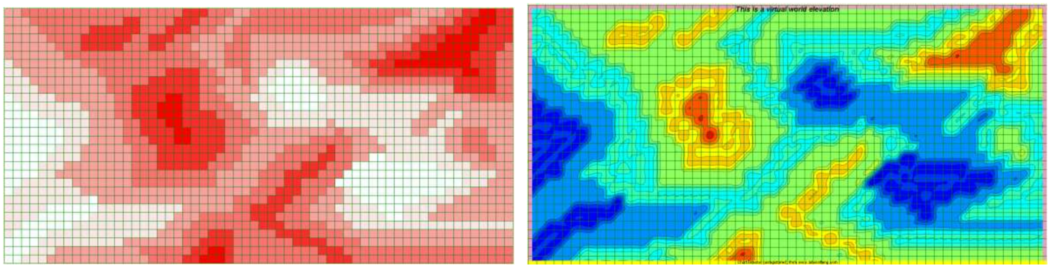


Figure 5.3: Topography representation of the study area

Left) using digital format method, Right) using contour maps

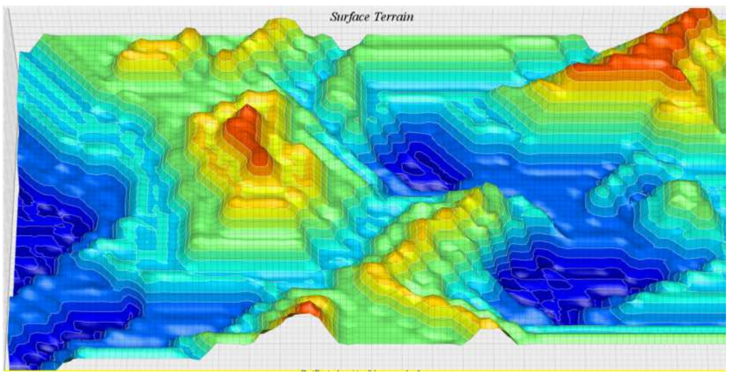


Figure 5.4: Surface terrain of a study area

5.3.2 THE COST COMPONENTS FOR A 3D HIGHWAY ALIGNMENT

As mentioned in section 4.3.2 of the previous chapter, several cost components can be considered for a highway alignment optimisation problem. In this study, the fitness function of 3D highway alignment was formulated by adding earthwork cost to the fitness function components of the 2D highway alignment described earlier. Earthwork cost is defined as the cost that the amount of cut and fill work incurs on the alignment.

Earthwork cost is affected by many factors. The amount of cut, the amount of fill, maximum grade limit, soil type, unit prices of cut and fill, hauling distances, and the weight that is being given to each of the two components during the optimisation process are among the most critical factors that affect earthwork cost.

In this chapter, the amount of earthwork in terms of cut and fill quantities was considered and added up to the fitness function components developed in Chapter 4. An approximate method was used to calculate the amount of cut and fill. The earthwork quantities were calculated without giving consideration to the side slopes. The difference in elevation between the alignment grade line and natural ground elevation for cut (h_c) and fill (h_f) between any two successive station points was calculated and directly multiplied by the length of that section (d_i), width of the road (w), and unit cost of cut (UCC) and/or fill (UCF). The cumulative earthwork of the consecutive sections would be the total earthwork amount. Figure 5.5 and Equations 5.5.1 to 5.5.7 illustrate the detail.

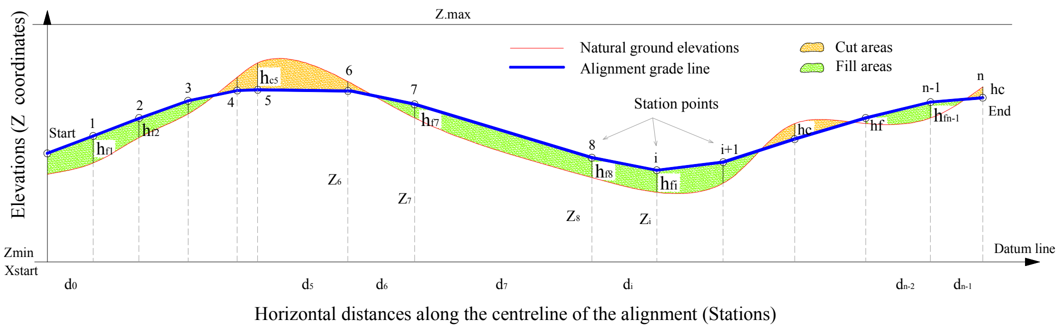


Figure 5.5: A sketch for earthwork explanation with station point concept

$$h_{c\_i_{av}} = \frac{h_{c_{i+1}} + h_{c_i}}{2} \tag{5.5.1}$$

$$h_{f\_i_{av}} = \frac{h_{f_{i+1}} + h_{f_i}}{2} \tag{5.5.2}$$

$$\mathrm{Q_{Cut}} = \sum_{i=0}^{n-1} \mathrm{d_i} * \mathrm{h_{c\_i_{av}}} * \mathrm{w} \tag{5.5.3}$$

$$Q_{Fill} = \sum_{i=0}^{n-1} d_i * h_{f\_i_{av}} * w \quad (5.5.4)$$

$$C_{Cut} = Q_{Cut} * UCC \quad (5.5.5)$$

$$C_{Fill} = Q_{Fill} * UFC \quad (5.5.6)$$

$$C_{Earthwork} = C_{Cut} + C_{Fill} \quad (5.5.7)$$

Where $C_{Earthwork}$ is the total earthwork cost, $h_{c\_i_{av}}$ and $h_{f\_i_{av}}$ are the average cut or fill depths between station points i+1 and i, respectively, while, d_i is the length of that section, and UFC and UCC are the unit fill and cut costs.

Thus, the total fitness function was to minimise:

$$C_{Total} = a_1.C_{Length} + a_2.C_{Location} + a_3.C_{Earthwork} \quad (5.6)$$

5.3.3 THE GA FORMULATIONS

5.3.3.1 Chromosome Representation

Using the ideas of station points, the method defines alignment through generating station points along the centreline of the alignment. The station points were considered as the decision variables of the alignment. Each decision variable (station point) in the chromosome was then considered as a gene encoded to X, Y, and Z coordinates.

For a 3D alignment, N number of station points was used to represent each chromosome as a candidate solution. The number of the station points, as with the 2D alignment, was highly related to the desired precision. Mountainous and hilly areas demand more station points than flat areas in order to provide smoother transitions between different grade lines at locations where they are traditionally known as sag and crest vertical curve locations (see Chapter 2 for details). For this purpose, the same Equations as 4.7.1, 4.7.2, and 4.7.3, described in Chapter 4, were used to determine an appropriate number of station points to represent the chromosome length.

The chromosome map was as shown in Figure 5.6; it contains the three-dimensional coordinates of each station point. The station points appear here in the order in which they occur along the length of the alignment. The order of any station point during its generation was kept until the end of the process and after finding the optimum solution. In other words, the points' order of the optimum solution was the same as when it was initially generated in the initial population.

Index (i)	0	1	2	…	i	…	N-1	N
Individual (j)	X_0	X_1	X_2		X_i		X_{N-1}	X_N
	Y_0	Y_1	Y_2		Y_i		Y_{N-1}	Y_N
	Z	Z_1	Z_2		Z_i		Z_{N-1}	Z_N

Figure 5.6: Genetic representation of a chromosome for a 3D highway alignment

5.3.3.2 Initial Population Generation

A random initial population of individuals was generated such that:

1. The station points' X and Y coordinates were within the study area.

$$X_{min} \leq X_i \leq X_{max} \text{ and } Y_{min} \leq Y_i \leq Y_{max} \tag{5.7}$$

Where

i is the index or series of the station points in the order they occur in the solution chromosome; *i= 1, 2, … , N-1*, and X_{min} and X_{max} are the lower and upper limits of the study area in the X direction, and Y_{min} and Y_{max} are the lower and upper limits in the Y direction. The coordinates of the station points must lie within these limits.

2. Two different initialisation methods were tested for the station points' Z coordinates (Elevations), as follows:

    - All station points' Z coordinates were within the maximum and minimum elevation bounds defined by the user, such that:

$$Z_{min} \leq Z_i \leq Z_{max} \qquad (5.8)$$

Where

i is the index or series of the station points in the order they occur in the solution chromosome; *i= 1, 2, … , N-1*, and

Z_{min} and Z_{max} are the lower and upper limits in the Z direction (elevation range). The coordinates of the station points must lie within these limits.

- All station points' Z coordinates were generated within a feasible bound specified by the limitations of the maximum gradient. The feasible limit of the search space of any station point (e.g; i) was determined by the maximum grade line limits drawn upward and downward from the two termini points and from the station point i-1, as follows:

To generate a Z coordinate within a feasible bound for the i^{th} station point, the horizontal distances (S_i) along the centreline of alignment and the elevations of the termini points should be known. Then follow these steps:

Step 1: Determine forward feasible bound (FZ_i^{upper} and FZ_i^{lower}):

$$FZ_i^{upper} = Z_{i-1} + \frac{S_i - S_{i-1}}{100} \; x \; G_{max} \qquad (5.8.1)$$

$$FZ_i^{lower} = Z_{i-1} - \frac{S_i - S_{i-1}}{100} \, x \, G_{max} \tag{5.8.2}$$

Where FZ_i^{upper} and FZ_i^{lower} are forward feasible upper and lower bound of the Z coordinate at station point i, respectively, specified by the maximum positive and negative grade lines drawn from station point i-1 (see Figure 5.7).

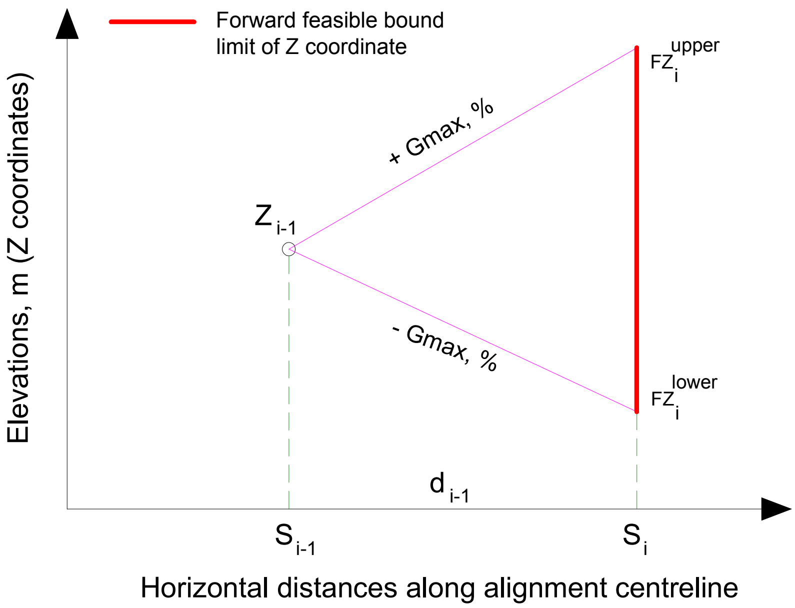


Figure 5.7: Forward feasible bound illustration for Z coordinates

Step 2: Determine backward feasible bounds (BZ_i^{upper} and BZ_i^{lower}):

$$BZ_i^{upper} = Z_n + \frac{S_n - S_i}{100} \, x \, G_{max} \tag{5.8.3}$$

$$BZ_i^{lower} = Z_n - \frac{S_n - S_i}{100} \, x \, G_{max} \tag{5.8.4}$$

Where BZ_i^{upper} and BZ_i^{lower} are backward feasible upper and lower bound of the Z coordinate at station point i specified by the maximum positive and negative grade lines drawn from the known end termini point (see Figure 5.8).

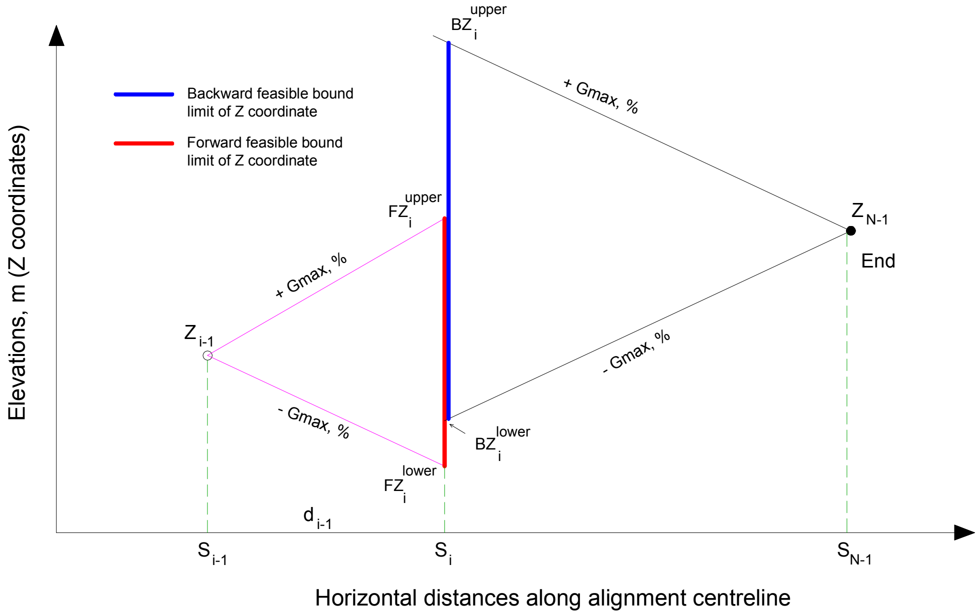


Figure 5.8: Backward feasible bound illustration for Z coordinates

Step 3: Determine the allowable Z_i^{upper} and Z_i^{lower} bounds:

$$Z_i^{upper} = \min\left(FZ_i^{upper}, BZ_i^{upper}\right) \tag{5.8.5}$$

$$Z_i^{lower} = \max\left(FZ_i^{lower}, BZ_i^{lower}\right) \tag{5.8.6}$$

Where Z_i^{upper} and Z_i^{lower} are the feasible bound of the Z coordinate at station point i.

Step 4: Generate random Z_i coordinates for the station point located at S_i distance from the start point, as follows:

$$Z_i = Rnd\ [Z_i^{upper}\ \ , Z_i^{lower}\ ] \tag{5.8.7}$$

so that ;

$$Z_i^{lower} \leq Z_i\ \leq Z_i^{upper} \tag{5.8.8}$$

This concept was first introduced by Fwa et al (2002) and used by Kang (2008). The whole concept is illustrated in Figure 5.9.

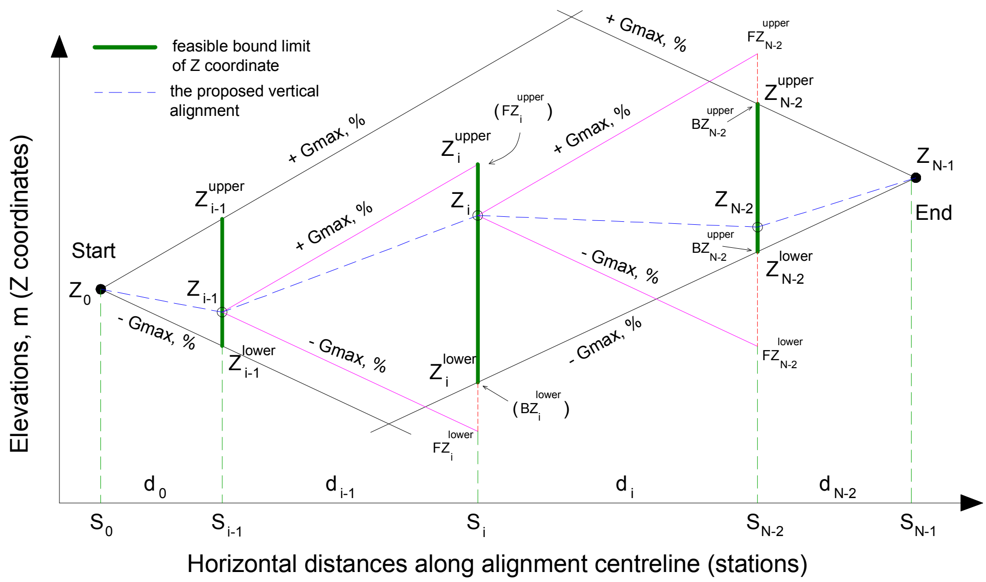


Figure 5.9: Z coordinate initialisation using feasible bound within the limits of G_{max}

3. The 3D components of each gene were encoded using floating point numbers.
4. The first and last points (0 and N-1) were fixed as the required terminal points.
5. The station points were sorted in the order of their X values ($X_i \leq X_{i+1}$). This process was specific to the initial population generation only and it was done just to produce initial candidates with less windings.

An initial vertical alignment solution that is randomly generated in a 3D area may look like the one shown in Figure 5.10.

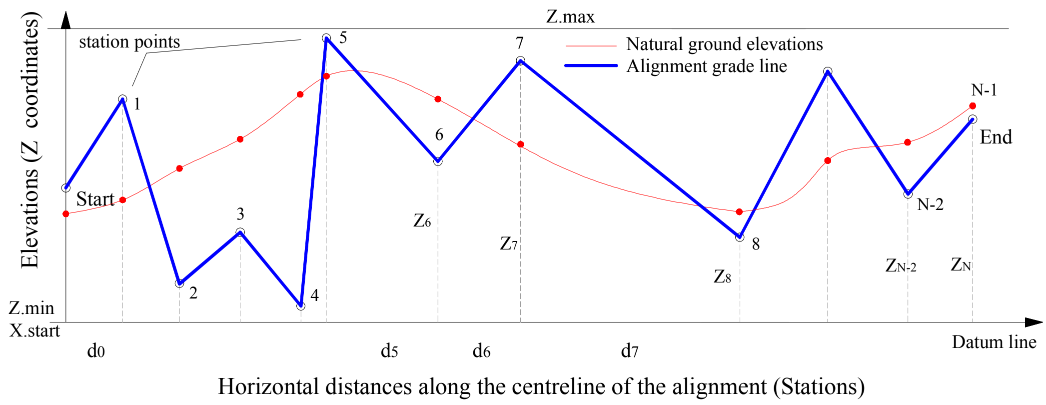


Figure 5.10: A randomly generated alignment profile in the initial population

5.3.3.3 Selection

i) Random-based Selection Scheme (RndSS)

In addition to the two selection methods described in Chapter 4, a third method called the Random-based Selection Scheme (RndSS) was also proposed. This operator selects the parent chromosomes for breeding in a random order. It was assumed that the design of this operator prevents incest and reduces incompatible mating, as happens with RnkSS and RevSS, respectively. The implications of using this method will be compared with RevSS and RnkSS used in Chapter 4. The parent population with this scheme, as for the other two selection schemes, is first evaluated and sorted according to their fitness value. Then, each two parents are randomly selected for breeding until the entire population is selected in such a way that no single candidate is chosen twice. The produced offspring population is recombined with the parents and sorted again and the best population is preserved for the next generation while the remaining population are killed off. Figure 5.11 shows the algorithm which implements RndSS:

For i = 0 to ($\frac{\text{Psize}}{2}$ – 1)   step 1

Generate two new random integers (e.g.; a and b) between (0 and P_{size} -1)

{Note: These two numbers must not be regenerated in the next steps}

Select Chr (a)

Select Chr (b)

Perform crossover

Next i

Figure 5.11: An illustration of RndSS

Where P_{size} is the population size, Chr denotes a chromosome. Also, note that zero (0) is also considered for counting the individuals in the population to match zero-based array performance in the programming language.

5.3.3.4 Crossover Operator

The Random multiple point crossover method described in Chapter 4 for 2D highway alignment, with up to eight crossover points, was used for 3D highway alignment optimisation. It should be mentioned that arithmetic crossover, which was introduced by Michalewicz (1999) to combine two chromosomes linearly, was also tested but showed no significant improvements upon the other methods.

5.3.3.5 Mutation Operator

The same uniform (SM) and grouped point mutation (GPM) methods described in Chapter 4 were used. The only difference was that the Z dimension was added to the formulations. Moreover, a combination of both methods was also tested: in addition to that, the mutation range was varied against time. These are described in the following sections (i-iii).

i) Standard GA Mutation Operator (Uniform Mutation, SM)

With this mutation, a position, say p, is randomly generated provided that $1 \leq p \leq N$. The decoded gene values (X, Y, and Z coordinates) at this position are replaced with new values within the limits of the specified boundary area:

Let X_p, Y_p, and Z_p be the coordinate values at position p before the standard mutation method is applied. Then, the operator assigns new values, as follows:

$$X_p^- = X_{min} + rnd\,\{X_{max} - X_{min}\} \quad (5.9.1)$$

$$Y_p^- = Y_{min} + rnd\,\{Y_{max} - Y_{min}\} \quad (5.9.2)$$

$$Z_p^- = Z_{min} + rnd\,\{Z_{max} - Z_{min}\} \quad (5.9.3)$$

ii) Grouped Point Mutation (GPM)

Using this mutation, new areas associated with alignment segments (represented by several station points), instead of a single point position as in the SM method, are explored. The mutation applies the same search strategy in the X, Y, and Z directions.

This method associates a number of sequential station points with a randomly selected and mutated point at R. The operator, as with simple uniform mutation, selects a gene position (R) randomly and assigns a new random value for its X, Y, and Z coordinates. Then, the operator generates two more locations (l_1 and l_2), provided that $l_1 < R < l_2$. All the genes (station points) that are located between l_1 and R on the left side of R, and l_2 and R on the other side are reallocated and put on a straight line connecting the newly generated gene at R with the selected genes at l_1 and l_2.

The mathematical steps of this mutation are detailed below:

1. Let A be an offspring chromosome before applying GPM, as shown in Figure 5.12.

Index	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	…	…	*N-1*
A chromosome before mutation															
Chromosome A	X_s	45	58	75	38	135	75	84	112	78	144	152	…	…	X_e
	Y_s	16	44	33	57	24	18	42	22	15	36	2	…	…	Y_e
	Z_s	250	253	258	235	255	260	262	250	252	255	251	…	…	Y_e

Figure 5.12: An offspring before GPM

2. Let R be a randomly generated position along the chromosome of length *N*.

$$R = rnd\ \{1, N-1\} \tag{5.10.1}$$

The selected gene at position R is decoded to the X, Y, and Z coordinate values.

3. Assign new values to the X, Y, and Z coordinates at R :

$$X_R^- = X_{min} + rnd\{X_{max} - X_{min}\}\ ; \quad where\ X_{min} \leq X_R \leq X_{max} \tag{5.10.2}$$

$$Y_R^- = Y_{min} + rnd\{Y_{max} - Y_{min}\}\ ; \quad where\ Y_{min} \leq Y_R \leq Y_{max} \tag{5.10.3}$$

$$Z_R^- = Z_{min} + rnd\{Z_{max} - Z_{min}\}\ ; \quad where\ Z_{min} \leq Z_R \leq Z_{max} \tag{5.10.4}$$

4. Generate two more random positions, l_1 and l_2:

$$l_1 = rnd\ \{1, R-1\} \tag{5.10.5}$$

$$l_2 = rnd\ \{R+1, N-1\} \tag{5.10.6}$$

Figure 5.13 shows the selection of the positions R, l_1, and l_2

Index	*0*	*1*	*2*	*l_1*	*4*	*R*	*6*	*7*	*8*	*l_2*	*10*	*11*	…	…	*N*
A chromosome before mutation															
Chromosome A	X_s	45	58	75	38	86	75	84	112	78	144	152	…	…	X_e
	Y_s	16	44	33	57	30	18	42	22	15	36	2	…	…	Y_e
	Z_s	250	253	258	235	255	260	262	250	252	255	251	…	…	Z_e

Figure 5.13: Position selections using GPM

5. Alter the location values of the points located between l_1 and R according to:

$$X_i = X^{-}_{R} - \left[\frac{X^{-}_{R} - X_{l1}}{R - l_1}\right] x\ (R - i) \qquad for\ all\ i = (l_1 + 1), \dots, (R - 1) \qquad (5.10.7)$$

$$Y_i = Y^{-}_{R} - \left[\frac{Y^{-}_{R} - Y_{l1}}{R - l_1}\right] x\ (R - i) \qquad for\ all\ i = (l_1 + 1), \dots, (R - 1) \qquad (5.10.8)$$

$$Z_i = Z^{-}_{R} - \left[\frac{Z^{-}_{R} - Z_{l1}}{R - l_1}\right] x\ (R - i) \qquad for\ all\ i = (l_1 + 1), \dots, (R - 1) \qquad (5.10.9)$$

6. Alter the values of the points located between R and l_2 according to:

$$X_i = X^{-}_{R} - \left[\frac{X^{-}_{R} - X_{l2}}{l_2 - R}\right] x\ (i - R) \qquad for\ all\ i = (R + 1), \dots, (l_2 - 1) \qquad (5.10.10)$$

$$Y_i = Y^{-}_{R} - \left[\frac{Y^{-}_{R} - Y_{l2}}{l_2 - R}\right] x\ (i - R) \qquad for\ all\ i = (R + 1), \dots, (l_2 - 1) \qquad (5.10.11)$$

$$Z_i = Z^{-}_{R} - \left[\frac{Z^{-}_{R} - Z_{l2}}{l_2 - R}\right] x\ (i - R) \qquad for\ all\ i = (R + 1), \dots, (l_2 - 1) \qquad (5.10.12)$$

The gene values of the resulting offspring after the application of this method would be as shown in Figure 5.14.

Index	*0*	*1*	*2*	*I_1*	*4*	*R*	*6*	*7*	*8*	*I_2*	*10*	*11*	…	…	*n*
A parent before mutation															
New Offspring	X_s	45	58	75	80.5	86	84	82	80	78	144	152	…	…	X_e
	Y_s	16	44	33	31.5	30	26.2	22.5	18.7	15	36	2	…	…	Y_e
	Z_s	250	253	258	256.5	255	254.2	253.5	252.7	252	255	251	…	…	Z_e

Figure 5.14: An offspring after GPM application

iii) Amendments to the mutation methods

With each of the above mutation methods (SM and GPM), different mutation ranges were applied to the single selected gene for mutation, as follows:

1. Random simple area consideration (RSA): This method applies simple random mutation within the boundary limitation of the study area, as described by Equations 5.9.1 to 5.9.3 and 5.10.2 to 5.10.4. This method allows free search within the boundary area and lets any part of the area be explored at any time.
2. Random dynamic area consideration (RDA): This method randomly reduces the search space limits in the X, Y, and Z directions. The aim behind this method is to produce various space limits at random (random small and big jumps). This operator has the ability to fine tune the solution on a random basis. This operator, for each selected individual to undergo mutation, generates a random dynamic area factor R_{area} so that $R_{area} \in \{0, \ldots, 1\}$, and is applied as follows:

Randomly generate a single precision value between 0 and 1 (R_{area}) and then find the search limits of X, Y, and Z, as below:

$$X_{upper\ limt} = X_R + (X_{max} - X_R) * R_{area} \quad (5.11.1)$$

$$X_{lower\ limt} = X_R - (X_R - X_{min}) * R_{area} \quad (5.11.2)$$

Randomly mutate the selected gene at R within the new defined limit:

$$X_R^- = X_{lower\ limit} + rnd\,\{X_{upper\ limit} - X_{lower\ limit}\} \tag{5.11.3}$$

These procedures are similarly repeated for Y and Z.

3. Uniform exponential area consideration (UEA): This method exploits the full bound of X, Y, and Z at the beginning of the search and exponentially reduces these bounds to reach zero at the end of the process. This method allows for a wide exploration within the whole area under consideration at the beginning of the search process and gradually starts to reduce and narrow the search bound until reaching zero at the end of the process. Narrowing the search limit by time will help to make a regular fine tuning of the solutions. The mechanism of this method is as follows:

   Let R_{area} be the area bound limit factor that is found as

$$\mathrm{R}_{area} = \left(1 - \frac{g}{G}\right)^{\left[1-\left(1-\frac{g}{G}\right)^{a}\right]^{b}} \tag{5.12}$$

Where

g is the current generation;

G is the total generation number;

a and *b* are user-defined parameters, and a = 1 and b = 1 are used here in the analytical tests.

The area ranges (area factor) that the above formula takes into account during the generation processes are as shown in Figure 5.15. The example is provided for total generation number G = 1000, a = 1, and b =1.

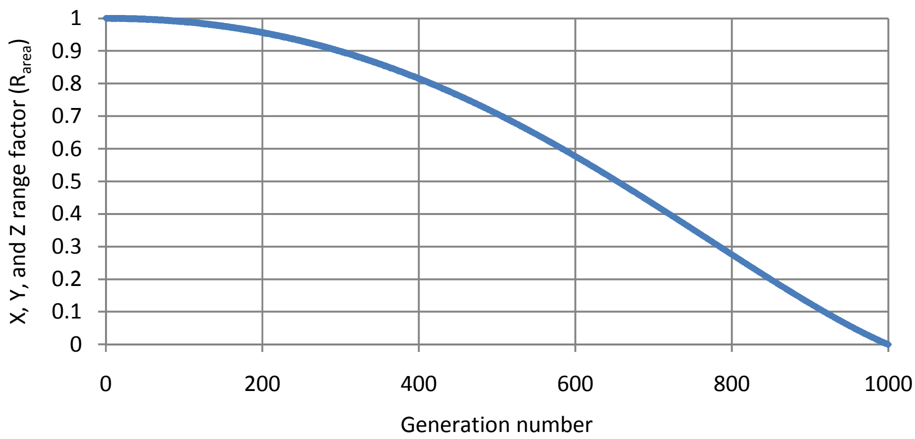


Figure 5.15: Mutation range factor by time

Then, applying the same formulas as for RDA to define the limits of X, Y, and Z, as follows:

$$X_{upper\ limt} = X_{Rm} + (X_{max} - X_{Rm}) * R_{area} \tag{5.12.1}$$

$$X_{lower\ limt} = X_{Rm} - (X_{min} - X_{Rm}) * R_{area} \tag{5.12.2}$$

Randomly mutate the selected gene at R within the new defined limit:

$$X_R^{-} = X_{lower\ limit} + rnd\ \{X_{upper\ limit} - X_{lower\ limit}\} \tag{5.12.3}$$

These procedures are repeated for Y and Z as well.

5.4 HIGHWAY VERTICAL ALIGNMENT CONSTRAINTS

The most critical constraints with highway vertical alignment geometric design are curvature and gradients. Both are important to produce a safe transition between any two points and to maintain the design speed. Details about vertical curvature and gradients as constraints are given in Chapter 6.

In this chapter, as the aim is to test the applicability of using the idea of station points for 3D highway alignment development, the bound of the Z coordinate (Z_{min}, Z_{max}) is therefore taken as the only relevant constraint to Z values in a 3D highway alignment.

5.5 EXPERIMENTAL TESTS AND THE RESULTS

A 3D highway alignment consists of a 2D alignment plus the third or Z dimension. Therefore, the 2D parameter result values that have been found (Chapter 4) were used as a base for the 3D highway alignment tests (Table 5.1).

Table 5.1: The result summary from 2D highway alignment model formulations

Pop_{size}	L_{chr}	CrO_Type	Mut-Method	%IMP	%PMR	Selection	Alignment feature
5000+	60	RMPCrO–8P	GPM (RDA)	5%-10%	3%-15%	RevSS	9 ± 1/10 within +97% of the most optimum. (for backward bends)
5000+	60	RMPCrO–8P	GPM (RDA)	5%-10%	8%-15%	RevSS	9 ± 1/10 within +97% of the most optimum. (for less bend requirements)
5000+	60	RMPCrO-8P	SM	5%-10%	3%	RevSS	9±1/10 within +60% of the most optimum solution

In the following sections, the possibility of building a model based on the proposed station point approach for 3D highway alignment development is investigated. The study starts with the model formulation verification and tests the applicability of the station point concept for 3D highway alignment optimisation. Next, different test scenarios are presented which aim to tune and test several combinations of parameters and methods. Table 5.3 abbreviates some 3D model-related expressions.

Table 5.2: Table of abbreviations for 3D modelling

Abbreviation	Description
W	World number
R_{width}	Road width in metre
UC_{cut}	Unit cost of cut (earthwork cut), unit cost/m^3
UC_{fill}	Unit cost of fill (earthwork fill), unit cost/m^3
Wt_L	Weighting factor that is given to location dependent cost
Wt_C	Weighting factor that is given to construction dependent cost (length dependent cost)
Wt_{EW}	Weighting factor that is given to earthwork dependent cost
C_{size}	Cell sizes of the area's rectangular grids, mxm
W_{width}	Width or length of the study area (designed world), m
W_{height}	Height of the study area (designed world), m
LUC	Location unit cost (Location dependent cost), unit/ml
UCC	Unit construction cost (length dependent cost), unit/ml

5.5.1 INITIAL TESTS AND MODEL VERIFICATION

This section aims to verify the model formulation and to test the applicability of the proposed station point approach for 3D highway alignment optimisation. Four different test scenarios were designed for comparisons (Table 5.3).

It is worth mentioning that these test scenarios share the same parameter value of population size (2000), length of chromosome (60), individual mutation rate (10%IMR), free random PMR, and location and length weighting factors equal to (1) and therefore they are not included within the parameter values listed in the table.

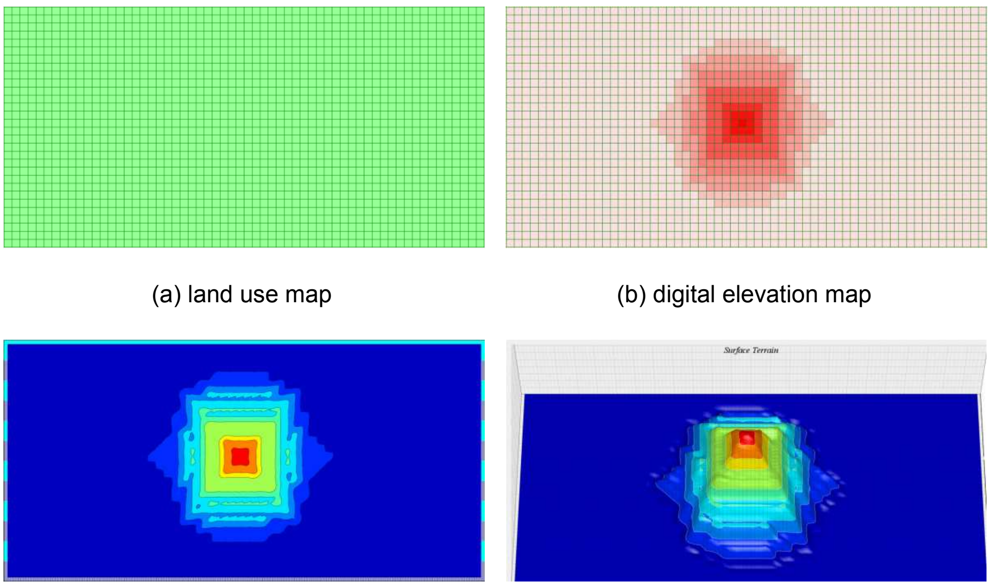
Table 5.3: Different test scenarios for 3D highway alignment model verification

Scenario	CrO_Type	MM	SS	Z consideration	G	Wt_L	Wt_{EW}
S1	RMPCrO - 8	GPM (RDA)	RevSS	Normal within Z_{min} and Z_{max} range	500	1	1
S2	RMPCrO - 8	GPM (RDA)	RevSS	Z Initialisation within feasible limit defined by G_{max}%	500	1	1
S3	RMPCrO - 8	GPM (RDA)	RevSS	Z Initialisation within feasible limit defined by G_{max}%	500	1	0.0003
S4	RMPCrO - 8	GPM (RDA)	RevSS	Z Initialisation within feasible limit defined by G_{max}%	500	1	0

For this purpose a specific world (W_8) was specially designed (Figure 5.16), with the following characteristics:

- the world was designed with unique land features of the same unit costs (unique location dependent cost) (see Figure 5.16 (a)),
- the world had the same unit construction cost, excluding the earthwork costs, per unit length of the alignment (construction dependent cost),
- the topography of the world consisted of a flat area (equal elevations) with a high land (mountain or hill) occupying the mid of the world (see Figure 5.16 (b-d)).

The aims of the tests, as mentioned in Table 5.3 above, were to verify the ability of the model to locate an optimum solution that skirts around the high elevation areas when equal weighting factors are used and also to verify the sensitivity of the model with different cost component weighting factors. The tests were also to compare and verify two mechanisms for Z coordinate initialisations (see Section 5.3.3.2).



(a) land use map (b) digital elevation map

(c) contour map (d) surface terrain

Figure 5.16: The land feature and topography of W_8

(Note: (a) is the land feature property which is made plain here, (b) is the digital elevation representation of the study area (the darker the cell the higher the elevation), (c) is the contour map (the dark blue colour is the lowest elevation and the red is the highest), and (d) is the surface terrain elevation of the area.) The world's parameter values are typed in Table 5.4.

Table 5.4: W_8 feature characteristics

W	W_{width} m	W_{height} m	C_{size} mxm	LUC unit/ml	UCC unit/ml	R_{width} m	UC_{cut} unit/m3	UC_{fill} unit/m3	Wt_L	Wt_C
W_8	12000	6000	200x200	100	300	9	10	10	1	1

Note that the tests were based on free selection of PMR. An explanation to this is that the land features of the areas were uniquely designed and configured and the elevations from start to end points around the hill were made equal. Thus, it was believed that free range of point selection (PMR) may produce straighter alignment segments quicker than if specific ranges were used.

The analyses of the tests were based on 10 test runs for each test scenario. Figure 5.17 shows the number of the solutions of those possessing fitness within specific fitness ranges of the most optimum solution. The minimum and maximum solution fitness of each scenario is presented in Figures 5.18 and 5.19.

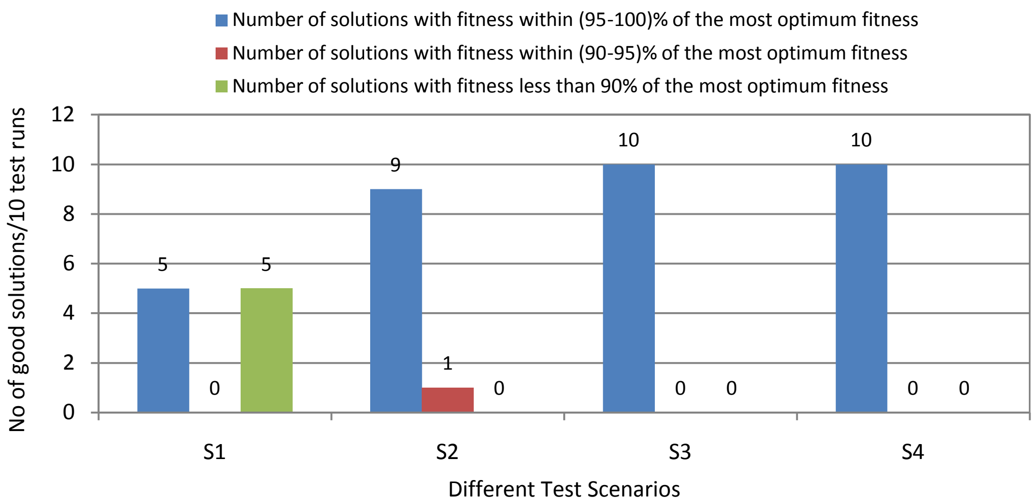


Figure 5.17: Number of good solutions obtained from different test scenarios

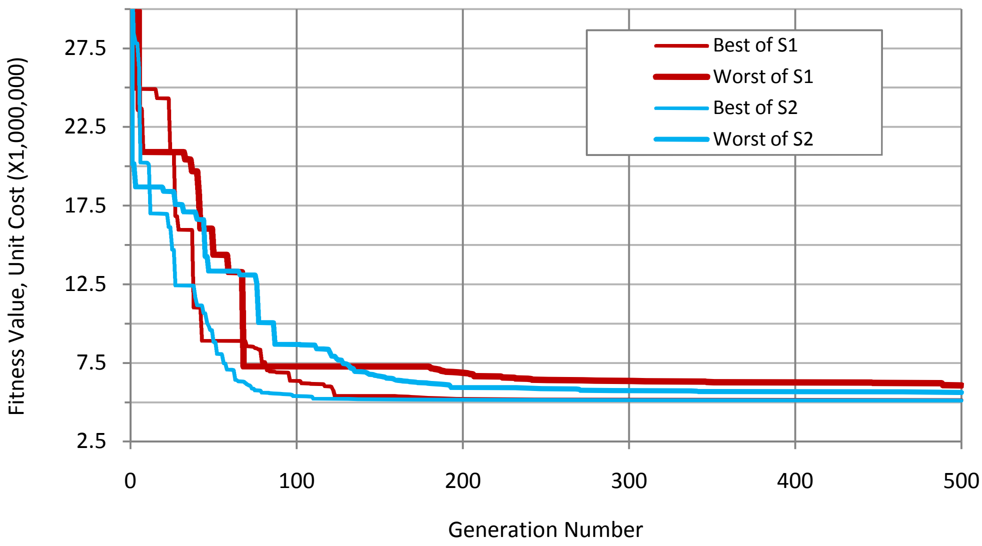


Figure 5.18: The effect of the Z initialisation method on solution convergence (S1 and S2)

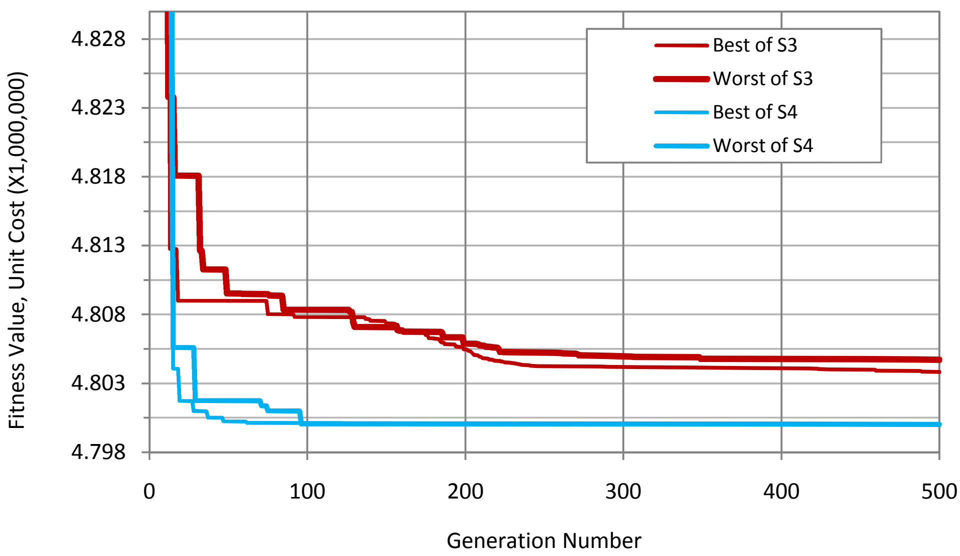


Figure 5.19: Sensitivity test of the model with different earthwork weighting factors (S3 and S4)

Note that the fitness result of scenarios 3 and 4 are incomparable due to the different cost weighting factors. They are put together to show how differently they behave and to prove that the model was sensitive towards different cost factors. This is clearly shown through the convergence of the solutions presented in Figure 5.19.

From the above test results the following conclusions could be drawn:

- The model formulations are correct
- The concept of station points for 3D highway alignment development is valid and further developments can be made
- Based on the designated world, the model is proved efficient in determining an optimum 3D highway alignment solution (see the optimum solution obtained from test scenario S2 in Figures 5.20 and 5.21)

  It can be noticed that the solution is straight at locations with the same land and elevation characteristics, skirts the high elevation areas, possesses zero grade, and it is put as close as possible to the natural ground elevation
- The model is proved sensitive with regard to different earthwork costs (compare the test results of scenario S2, S3, and S4) (see also Figures 5.22 – 5.24)
- It is also proved that Z initialisation by using the feasible bounds defined by G_{max} is more efficient than using Z_{max} and Z_{min} bounds (compare the results from scenario S1 and S2 and for more details see section 5.3.3.2). Figures 5.25 and 5.26 show the difference between the two initialisation methods, which show two best solutions in the initial population generated by each method. The range of Z_{min} and Z_{max} for Z coordinate initialisation (Figure 5.25) is varied between 530m and 720m, respectively.

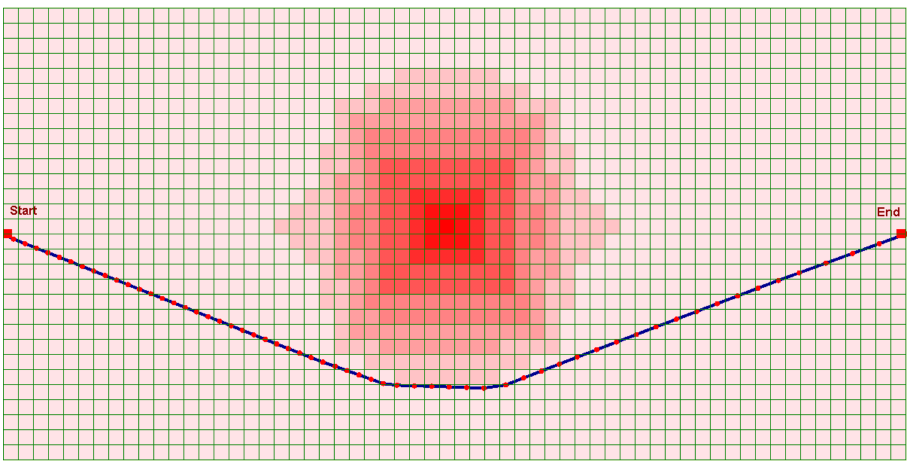


Figure 5.20: The optimum horizontal alignment obtained with test scenario S2

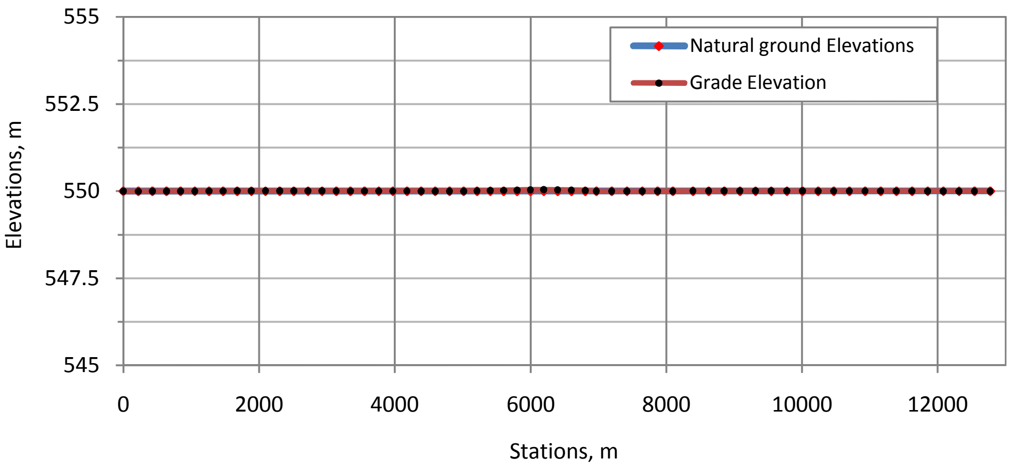


Figure 5.21: The optimum vertical alignment obtained with test scenario S2

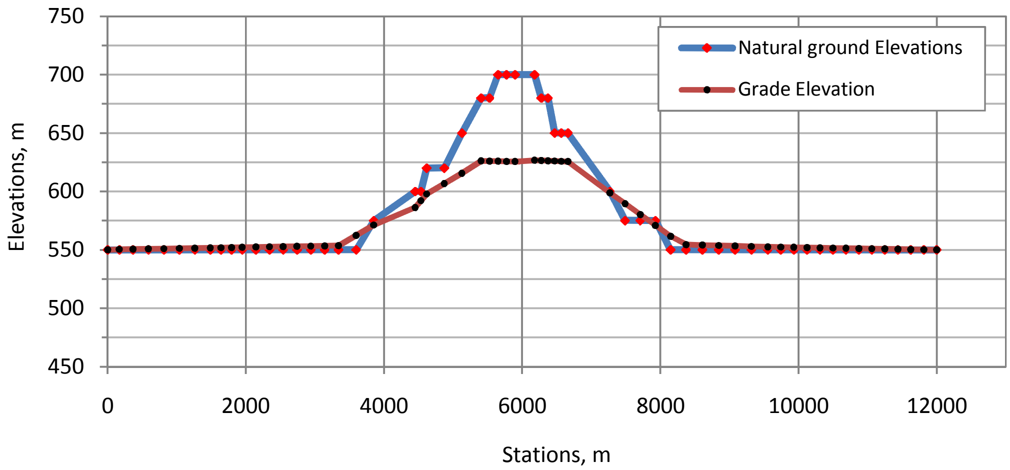


Figure 5.22: The optimum vertical alignment obtained using test scenario S3

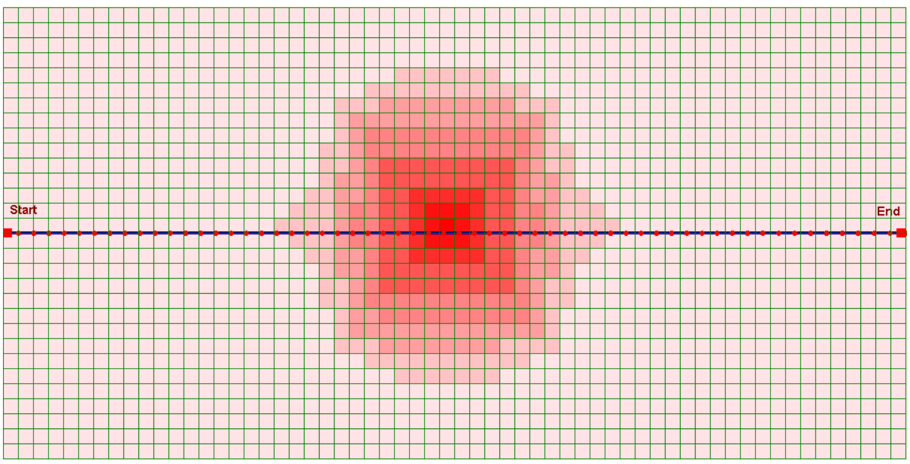


Figure 5.23: The optimum horizontal alignment obtained with test scenario S4

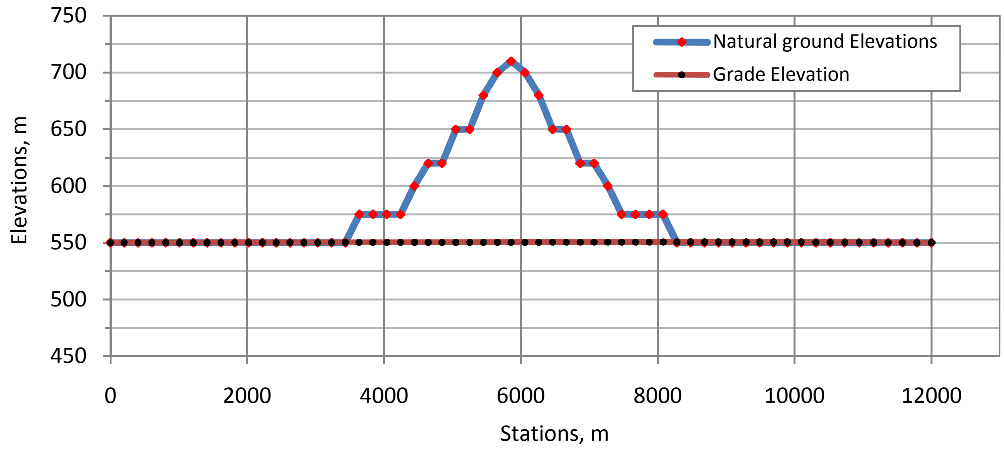


Figure 5.24: The optimum vertical alignment obtained with test scenario S4

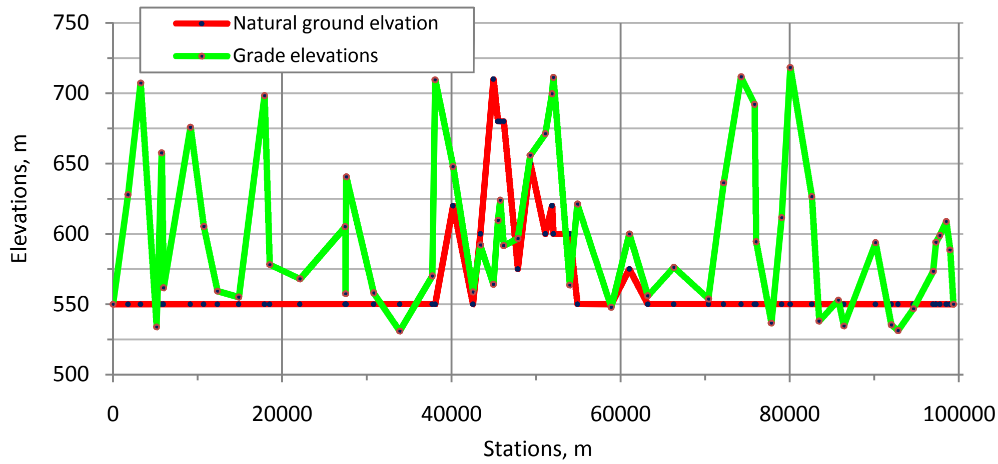


Figure 5.25: An initial alignment profile generated with using the standard Z initialisation technique within Zmin and Zmax limits (scenario S1)

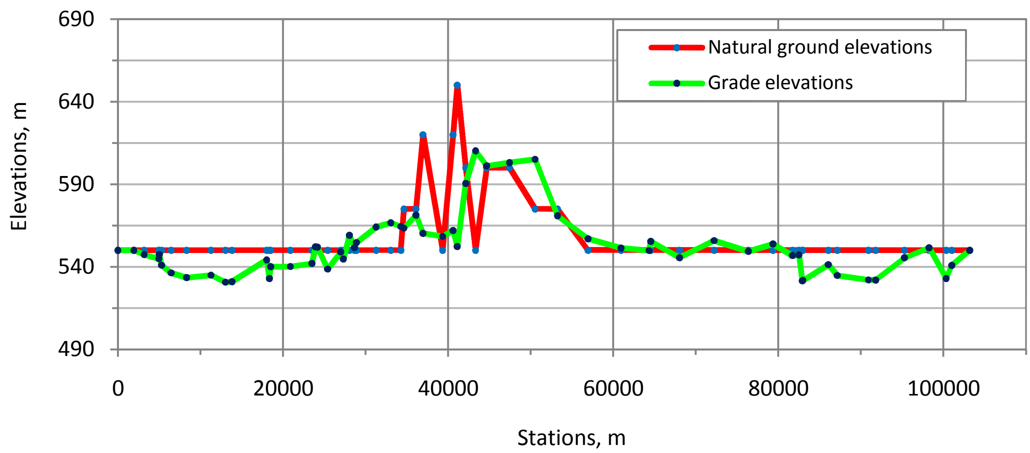


Figure 5.26: An initial alignment profile generated with using feasible Z bounds defined by limits of G_{max} (S2)

From the above test results, it can be concluded that the model performs quite well in recognising the areas that incur least earthwork costs. Moreover, the model has simultaneously dealt with horizontal and vertical alignments and produced optimum 2D and 3D alignments. Thus, the combination of the horizontal and vertical alignments was a solution with least possible fitness value. It is also worth mentioning that the model behaved differently with different test scenarios, whereby the sensitivity of the model was confirmed.

5.5.2 SECOND STAGE TESTS

This stage of testing aimed to test the model on a more complex world. For this purpose several test scenarios were made to test some methods and tune some parameters associated with the model for 3D highway alignments. To do this, a world W_9 was specially designed in such a way that the different land and elevation features would match some features that could exist in nature. Imitating real world features for testing the model increases the reliability of the model in real world applications. This world was generated in three stages. First, the land features were randomly generated; second, some of the land features were manually edited to match real world cases; third, topographical elevations were manually attached to the world in such a way that ensures providing different elevation configuration and matches the land configurations in the 2D plane. This would test how both horizontal and vertical alignments could be at the optimum simultaneously. The land and topographic features of this world (W-9) are presented in Figure 5.27.

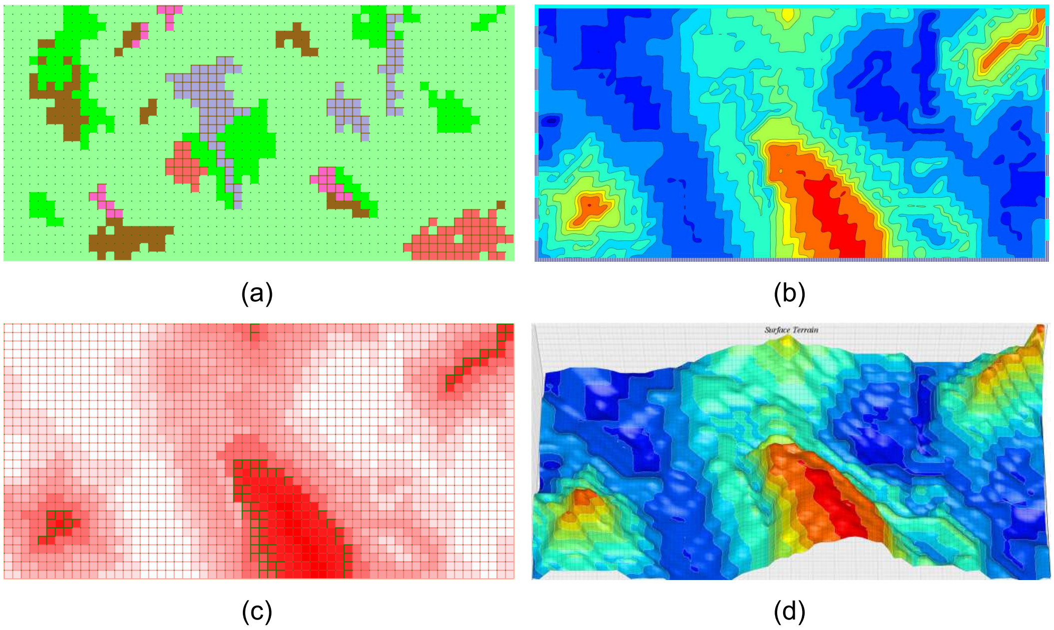


Figure 5.27: W_9 configurations, (a) 2D land features, (b) Contour map (the dark blue colour is the lowest elevation and the red is the highest), (c) Digital elevation representation of the study area (the darker the cell, the higher the elevation), (d) 3D surface terrain of the study area

The colour gradient definition of the land and topographical features are explained in Figure 5.28.

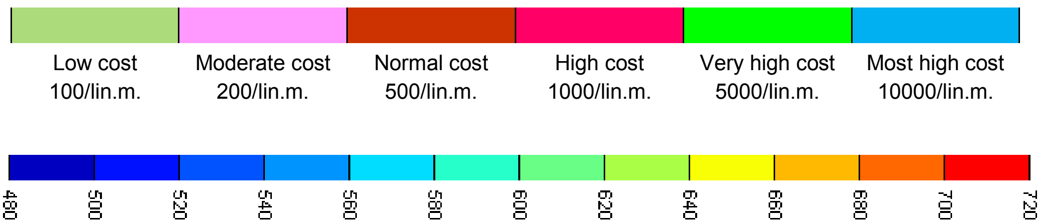


Figure 5.28: Upper colour gradient bar: different 2D land features of different unit costs;

Lower colour gradient bar: elevation gradient (m)

The tests were conducted using the test parameters that led to producing the optimum or near optimum 2D highway alignment solution. Moreover, some other test scenarios with different methods and parameters were also designed and the results were compared. It should be noticed that the test scenarios were made different through using different mutation, selection, and Z coordinate initialisation methods. Table 5.5 lists the test scenarios that were applied on W_9 to optimise a 3D highway alignment.

Table 5.5: Different test scenarios for world W_9

Scenario	CrO_Type	MM	SS	Z consideration	Wt_{EW}
S1	RMPCrO-8P	GPM (10% PMR) (RDA)	RndSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S2	RMPCrO-8P	GPM (10% PMR (RDA)	RnkSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S3	RMPCrO-8P	GPM (10% PMR) (RDA)	RevSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S4	RMPCrO-8P	GPM (10% PMR) (RSA)	RevSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S5	RMPCrO-8P	GPM (10% PMR) (UEA: 1/1)	RevSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S6	RMPCrO-8P	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	RevSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S7	RMPCrO-8P	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	RndSS	Z Initialisation only within feasible limit defined by G_{max}%	1
S8	RMPCrO-8P	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	RnkSS	Z Initialisation only within feasible limit defined by G_{max}%	1

(*) The exponential powers a and b are equal to 1 and 1 respectively (see Equ. 5.13).

Figure 5.29 shows the number of good solutions per 10 test runs. The numbers of good solutions were counted based on the percentage of the solution fitness values to the fitness of the most optimum solution. Here, a solution was considered the most optimum when it possesses the least fitness among the entire test scenarios and runs. This graph compares the stableness of the test parameters in producing solutions closer to the most optimum one. The higher the number of good solutions that are closer to the most optimum fitness, the more stable and reliable is the test scenario.

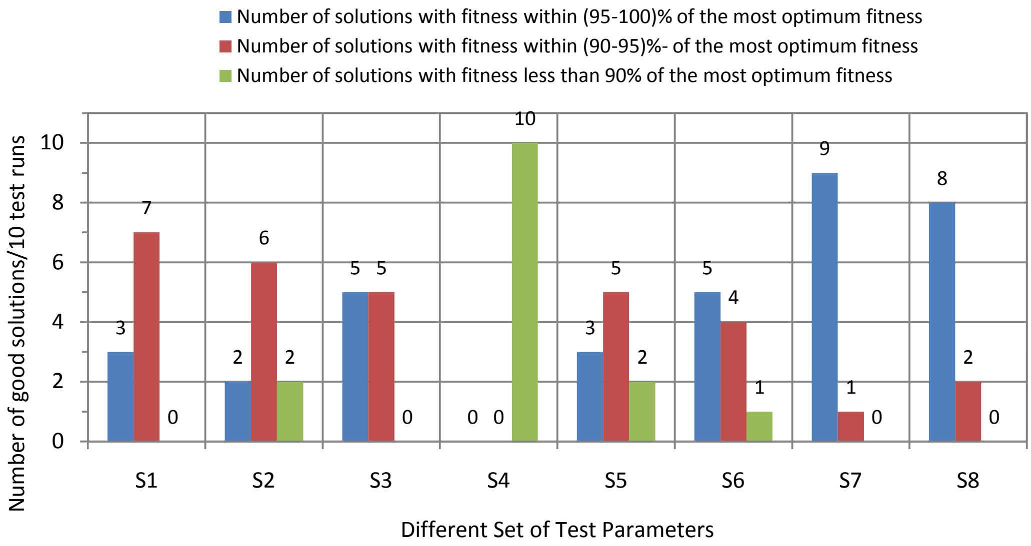


Figure 5.29: The effect of different test scenarios on obtaining stable results

Figures 5.30 and 5.31 show the range of minimum and maximum fitness of each scenario. These graphs also compare the convergence behaviour of the test scenarios at the successive generations. The minimum fitness graph shows how efficient are the different parameters in producing optimum or near optimum solutions. Moreover, the maximum fitness graph shows how far a solution would be from the optimum using the specified test parameters. Figures 5.32 to 5.35 present the same information but in different scales.

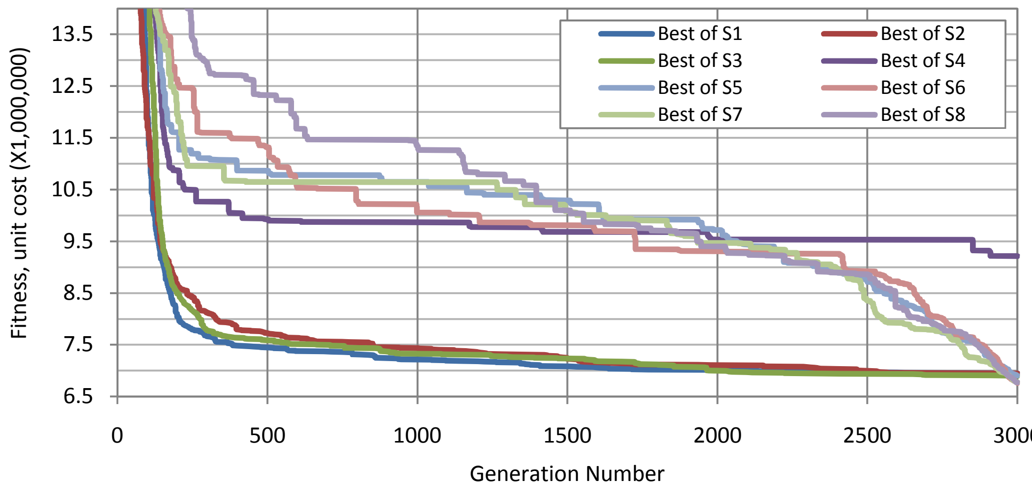


Figure 5.30: The effect of different test scenarios on solution convergence behaviour and minimum fitness value from 10 test runs/scenario

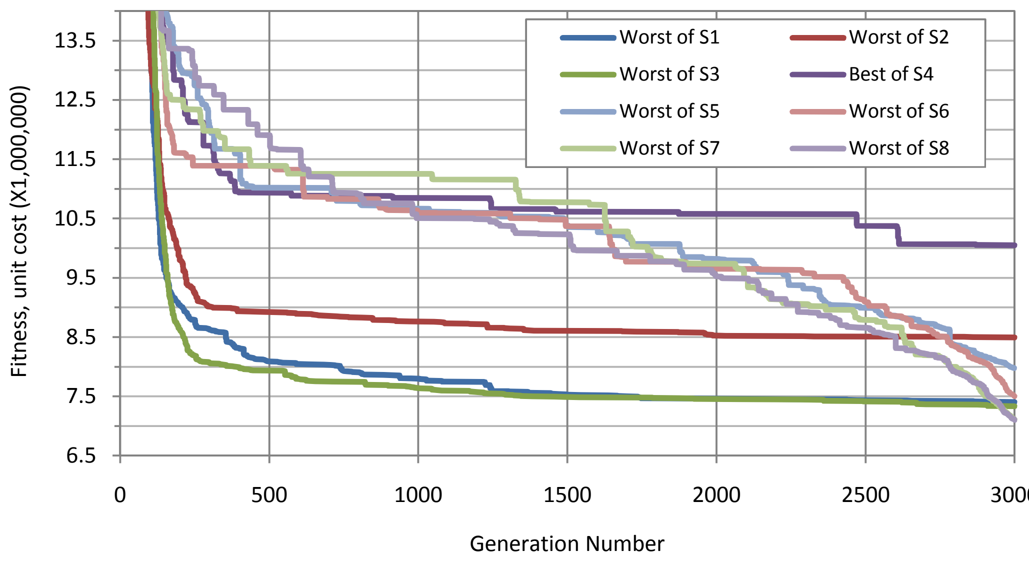


Figure 5.31: The effect of different test scenarios on solution convergence behaviour and maximum fitness value from 10 test runs/scenario

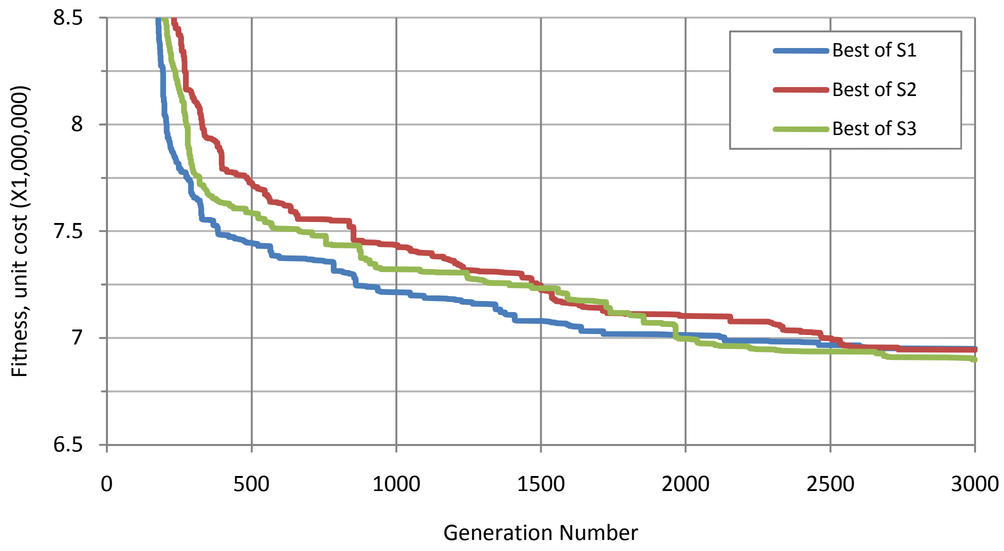


Figure 5.32: Minimum fitness of several scenarios from Figure 5.30 with a different scale

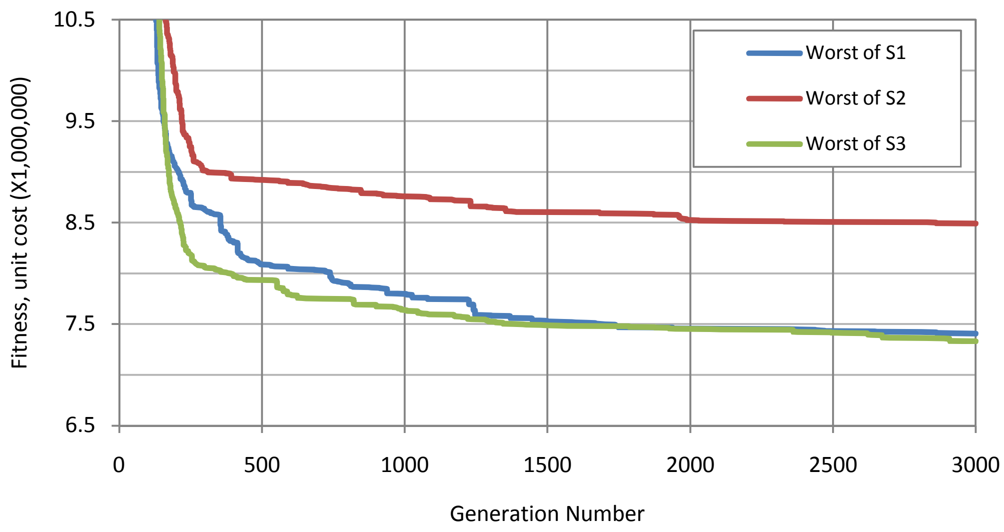


Figure 5.33: Maximum fitness of several scenarios from Figure 5.31 with a different scale

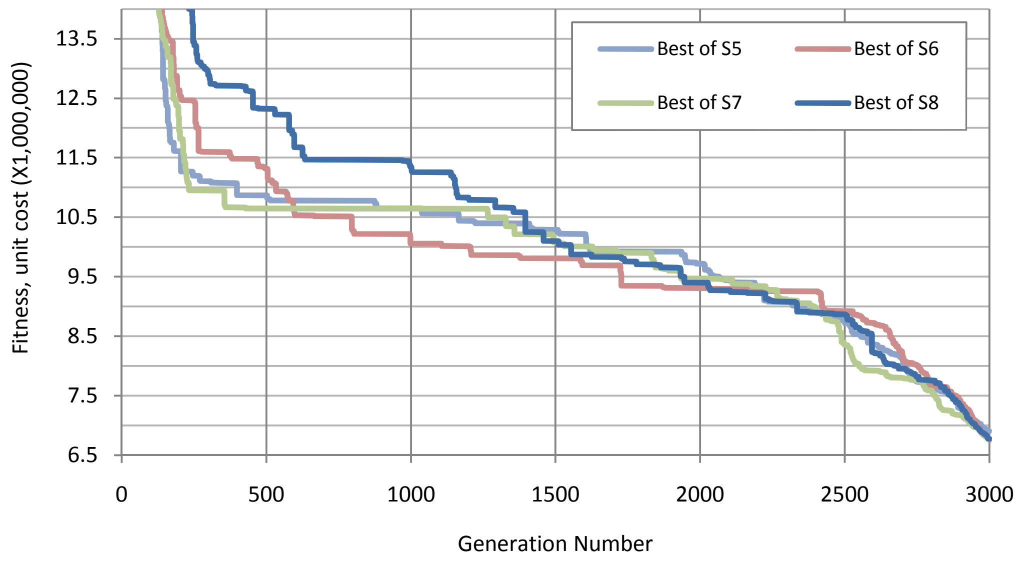


Figure 5.34: Minimum fitness of several scenarios from Figure 5.30 with a different scale

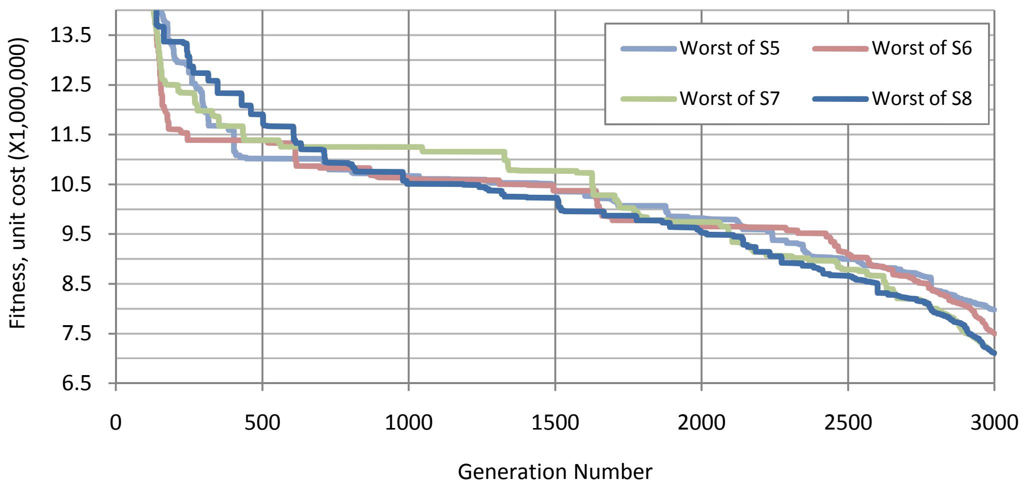


Figure 5.35: Maximum fitness of several scenarios from Figure 5.31 with a different scale

Figure 5.36 shows the percentages of the minimum and maximum solution fitness values obtained to the most optimum one.

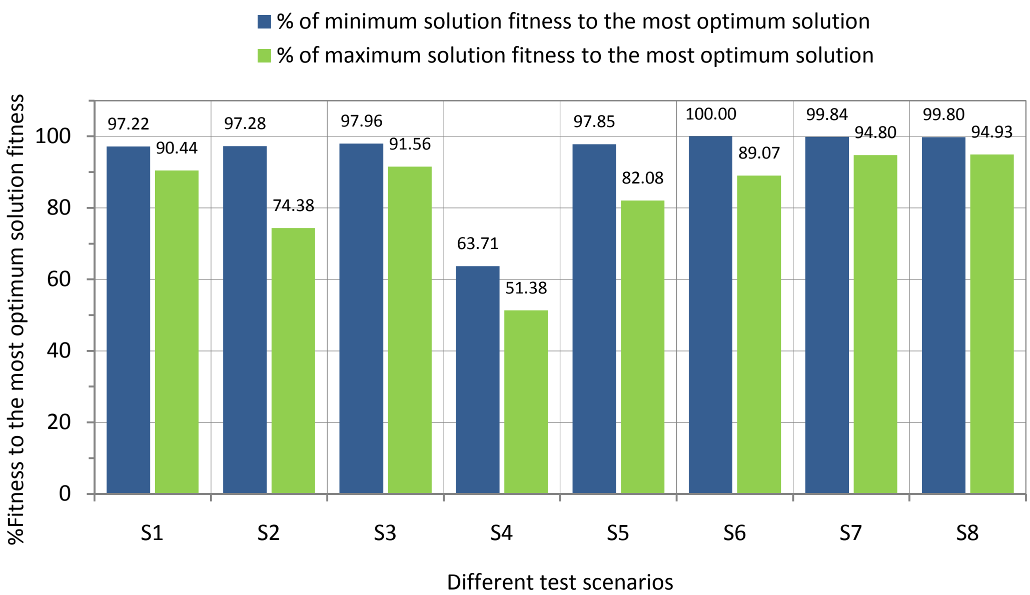


Figure 5.36: Percentage of minimum and maximum solution fitness per test scenario to the most optimum solution fitness

A decision on the best performance from the above test results is difficult. A user should compromise among several requirements and cannot depend on one graph presentation. For instance; when the evaluation is made depending on the most stable solution, test scenarios 7 and 8 are the best. When the minimum fitness is considered, test scenarios 6, 7, and 8 would be the best. Test scenarios 8 and 7 can also be seen as the best with regard to the highest solution fitness as they produced solutions with lower fitness variances. Therefore, and from the viewpoint of a stable test scenario and minimum fitness value, the test scenarios 6, 7, and 8 can be seen as the most efficient scenarios.

The results divided the test scenarios into two groups. A group, let it be G1, which was defined by the test scenarios 1, 2, 3, and 4, behaved quite differently than the other group, let it be G2, which was defined by sets 5, 6, 7, and 8. G1 converges much faster than G2 while G2 performed better. With G1, efficient search was performed within the first 500 generations and the population was almost converged at generation 1,500. Comparing the best of G1 (S3) with the best of G2 (S6) at generation 1,500, the fitness were 7,232,652 and 9,807,353, respectively, whereas the result with S6 was better than

S3 at the end of the search process, where the fitness were 6,760,786 and 6,898,927, respectively.

The difference in behaviour of the two groups can be explained by the way that the mutation is applied. With G2, the search starts by exploiting the entire search space (global search) and exponentially starts to reduce the area of search (local search) to reach zero at generation G (G is the total generation number). Using a wide range search at the beginning of the process helps a global exploration, while the narrow range search at the end helps to perform local searches and fine tune the solution.

Figures 5.37 to 5.39 show an optimised horizontal and vertical alignment using test S7. The solution has a fitness value equal to 6,771,815 (which is fit as 99.84% as the most optimum solution) with earthwork quantities of 23,192 m^3 and 44,431 m^3 for cuts and fills, respectively. The resulting alignment, due to imposing no constraints for a vertical grade limit, possessed a maximum grade of -118.29%.

From Figures 5.37 and 5.38, it can be seen that the alignment passes through the low unit cost land features and skirts the high elevation areas. Moreover, the vertical alignment (Figure 5.39) was put fairly close to the natural ground elevation to reduce and minimise the amount of cut and fill. Sharp grades up and downwards can be noticed from Figure 5.39. This was because gradients, as constraints, were omitted from the formulations of this stage, which will be addressed in the next chapter. The horizontal and vertical alignment results revealed the efficiency of the model to simultaneously produce a 3D solution with the least possible fitness value.

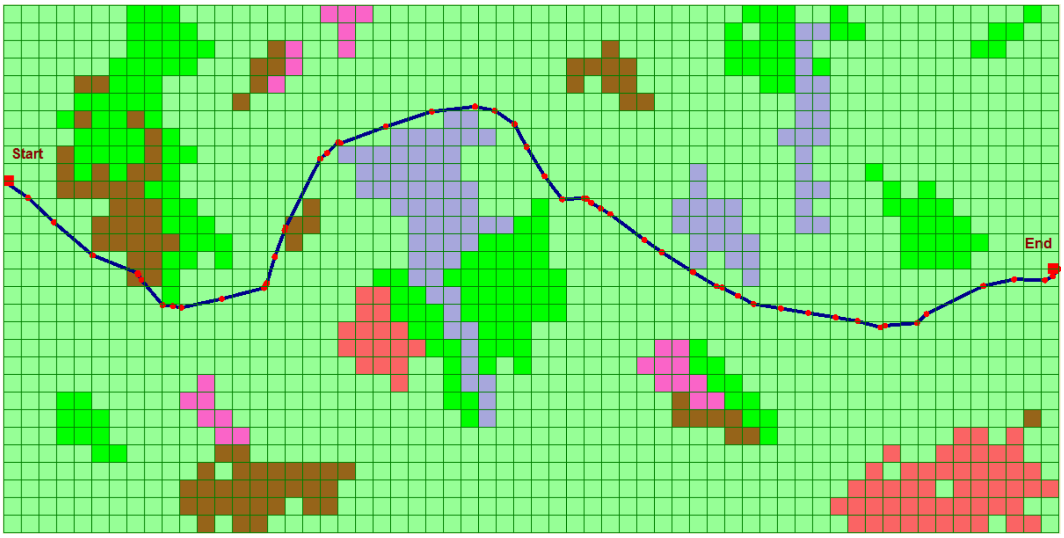


Figure 5.37: A 2D optimum horizontal alignment solution on land features' map at generation 3000 generated with test S7

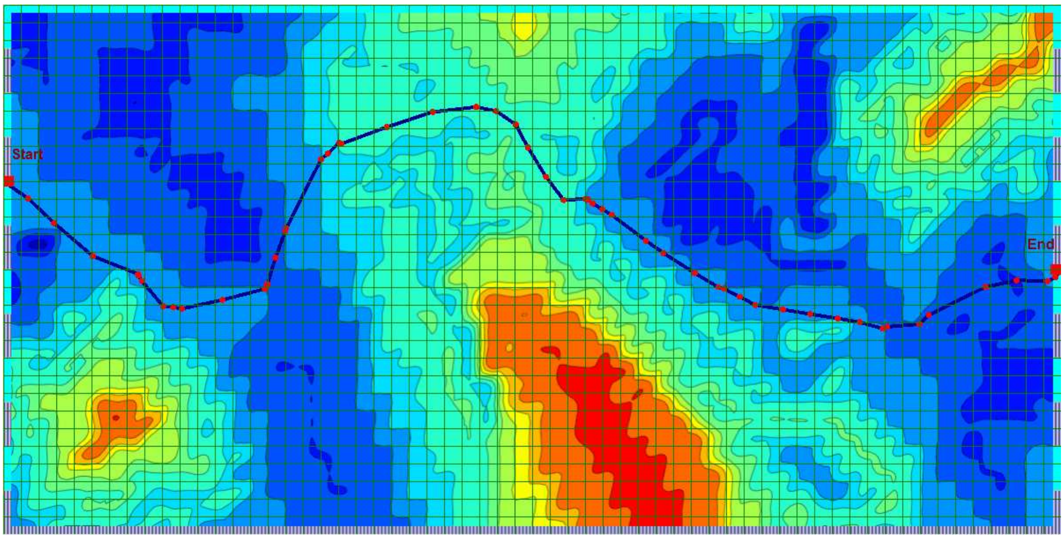


Figure 5.38: A 2D optimum horizontal alignment solution on contour map at generation 3000 generated with test S7

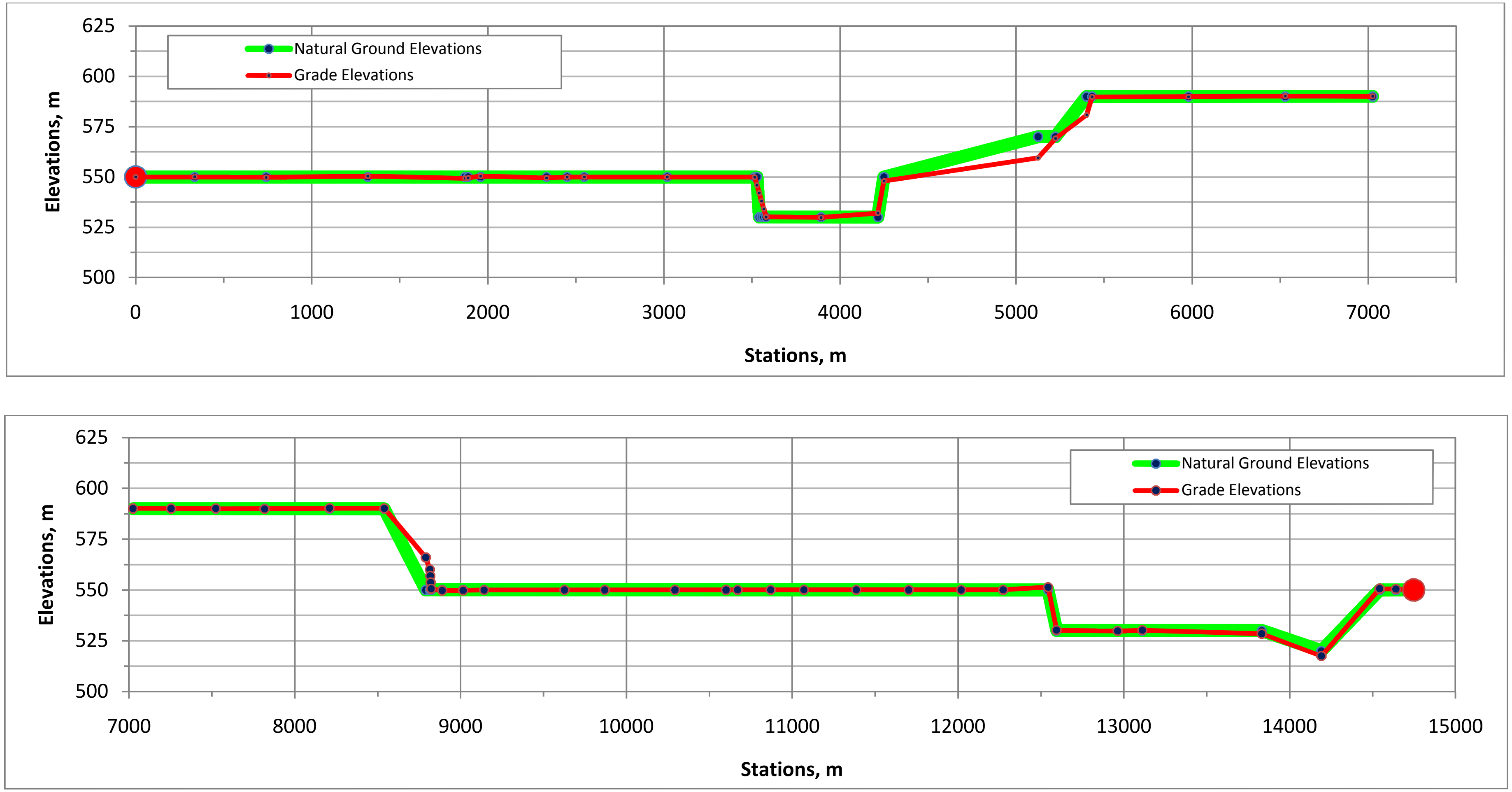


Figure 5.39: The profile (grade line) of the optimum solution at generation 3000 from test S7

A verification test was performed to check whether a simultaneous horizontal and vertical optimisation is necessary or not to produce a global 3D optimum solution, and also to show how the model behaves with and without consideration of the third dimension (Z coordinates). For this purpose, the same parameter values of S7 were used to generate a 2D highway alignment using the same world W_9. Ten test runs were conducted and the fitness results produced 10 good results out of the 10 test runs, in which they possessed fitness values within 98%-100% of the most optimum solution. The optimum result was as shown in Figure 5.40.

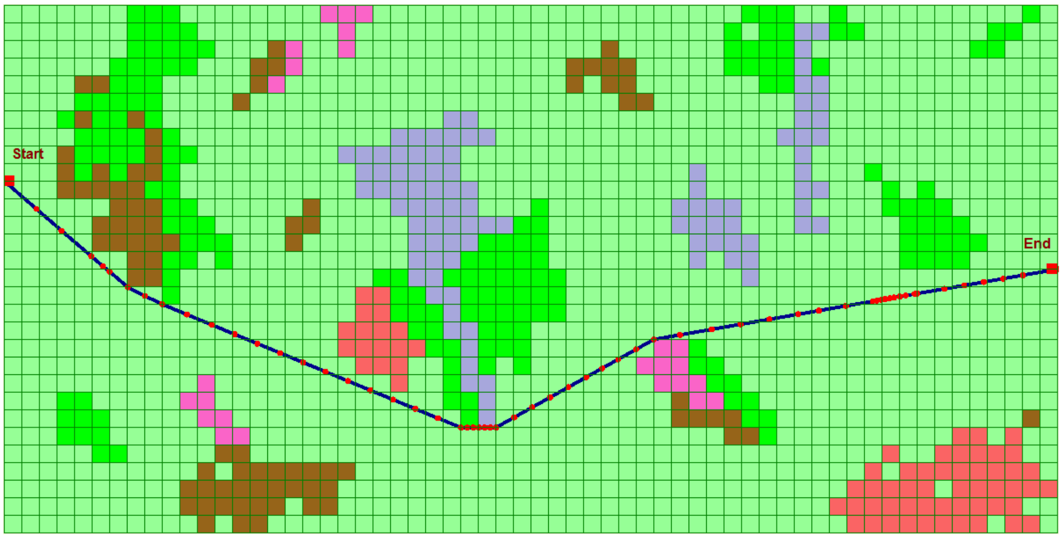


Figure 5.40: A 2D highway alignment result with no Z (third dimension) consideration at generation 3000 generated with test S7

The configuration and location of the global optimum 2D solution (Figure 5.40) was far from the global optimum 3D solution presented in Figure 5.37. This reflects the effect of the third dimension on the final location and configuration of the optimum solution. Therefore, it is strongly recommended to consider both horizontal and vertical alignment simultaneously during highway alignment optimisation processes; otherwise, the obtained solution would be locally optimum rather than being global.

It is also worth mentioning that the alignment configuration resulting from both models to generate 2D and 3D alignments has rough configurations for both horizontal and vertical alignments. Moreover, the grades of some of the line segments along the alignment grade line (profile) have very big (steep) values which make the alignment tougher and not realistic according to the limitations defined by the standard specifications. Producing a highway alignment solution that meets the requirement for horizontal and vertical curvatures, in addition to gradient constraint, are explored and investigated in the next chapter.

5.6 SUMMARY

A model for 3D highway alignment optimisation was developed. The aim was to investigate the possibility of using station points for 3D highway alignment development and to test the possibility of considering both horizontal and vertical alignments simultaneously. Simultaneous horizontal and vertical alignment consideration was important to obtain a global solution rather than a sub-optimal one, as is the case with the traditional (existing) design method. The usability of GA for the 3D problem was also evaluated through the implementation of the model. Finding optimum positions for the station points in a 3D space was required to configure an optimum 3D alignment solution. For these purposes several test scenarios were designed to tune the model parameters that can generate optimum alignment solutions.

A 3D world scenario with relatively complex land and topographic features was designed with which the model parameters and algorithms were tested and tuned upon. The accuracy of the solutions obtained, as with the 2D model described in Chapter 4, was constrained by the grid cell sizes and the number of station points used to represent the study area and the alignment, respectively. A rough 3D alignment was sought through connecting the station points using straight line segments in a 3D space. The study area was defined as a rectangle with a precision defined by the grid cell sizes. It was assumed that the location and topographic elevations within the individual grid cells were uniform.

The objective function was formulated to deal with location, construction, and earthwork dependent costs. Using this function, in addition of avoiding high cost fields and producing solutions with fewer lengths, as was the case with the 2D model presented in Chapter 4, the amount of earthwork was also minimised. The model is unable to deal with mass haul costs including the requirement for different cut and fill costs within the study area. These issues are recommended for future investigations.

The model was formulated so that 3D highway alignments are developed through position definition (X, Y, and Z coordinates) of the station points. The solutions obtained by the model could not be more than a piecewise linear trajectory in a 3D space. Apart from the boundary of the study area in the X, Y, and Z directions no specific constraints were imposed on the solution algorithms for realistic alignment generation.

The results showed how efficient is the model utilising the notion of the station points for 3D highway alignment development. It was found that the proposed design approach of station points are valid for 3D highway alignment configuration besides showing that horizontal and vertical alignments could not be optimised in two independent stages. The results indicate that simultaneous horizontal and vertical alignments are inevitable for a global optimum solution. The GA technique as a search tool was also proved successful to deal with the problem.

It was seen that no consideration was given to the smoothness of the horizontal and vertical alignments, as well as the vertical gradients. An alignment without the satisfaction of horizontal curvature, vertical curvature, and vertical gradients cannot be regarded as a realistic alignment. These constraints are important for safe driving conditions; the next chapter is dedicated to dealing with these issues.

CHAPTER 6

CONSTRAINT HANDLING ALGORITHMS AND 3D HIGHWAY ALIGNMENT MODEL

6.1 INTRODUCTION

The concept of station points for highway alignment development has been shown to be successful based on the experiments presented in Chapters 4 and 5. A 3D highway alignment can now be optimised using the model introduced in the previous chapters. However, the alignments that are obtained by the model cannot be considered as realistic as they do not obey the safety limitations defined by standard design requirements. The model so far is only able to produce a rough alignment with sharp horizontal and vertical curvatures alongside steep grades that occur at many locations along the alignment. Moreover, a solution cannot be regarded as good unless the main standard limitations are included.

This chapter is about handling the constraints that are important to produce a smooth and realistic 3D highway alignment. The constraints that are considered here are horizontal curvature, vertical curvature, and maximum vertical gradient. Furthermore, balancing the requirement for both cut and fill quantities are also discussed. The viability of handling these constraints with the concept of station points is the main aim driving the formulations observed in this chapter. Different algorithms and formulae are suggested and developed to deal with these constraints.

This chapter also presents several scenarios used to test the sensitivity of the model under different criteria. At the end of the chapter a verification test is conducted using a different world to ensure that the model is efficient under different circumstances by which the model reliability in real world applications is increased.

6.2 THE HORIZONTAL CURVATURE AS A CONSTRAINT

Horizontal curvature is traditionally defined as the degree of the curve. Degree of the curve is a function of the curve radius and the design speed (see Section 2.4.1, Chapter 2). Although most of the traditional equations that are used for highway geometric design elements and curvature definition cannot be used here with the concept of station points, governing principles and standard requirements for smooth horizontal transition and safe driving condition still need to be satisfied.

Recalling the traditional principles for horizontal curvature and by using the chord definition as the degree of the curve (D_c), which is defined by the angle subtending a 100ft (30.5m) chord length, the transition in the horizontal direction between any two successive chords is created by a deflection angle not greater than D_c, which is found using equation 6.1. The general form and the principle of chord definition for the degree of the curve are as shown in Figure 6.1. Ideally, the allowable degree of the curve (e.g. allowable Dc.(i)) is the maximum allowable deflection that can be produced between any two successive chords (e.g. $L_{chord.(i-1)}\ and\ L_{chord.(i)}$) to make a change in horizontal direction of which both a smooth and a safe driving condition are obtained.

$$D_c = 2 * \sin^{-1}\left(\frac{30.5/2}{R_{min}}\right) \tag{6.1}$$

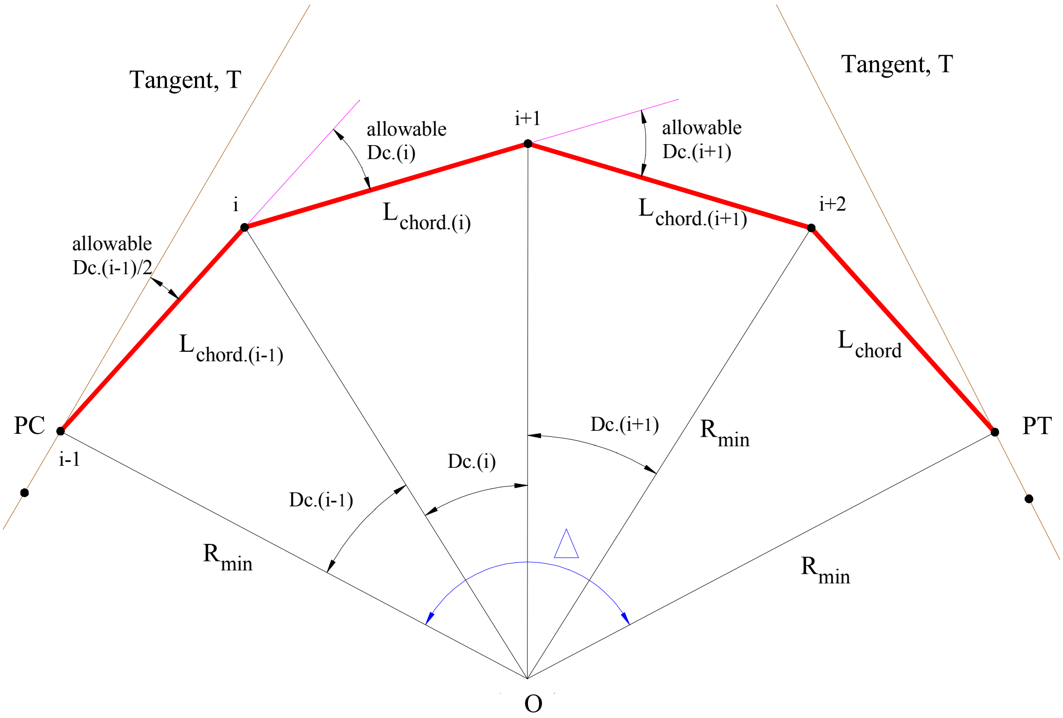


Figure 6.1: Chord definition illustration as degree of the curve

From Figure 6.1 and based on trigonometric identities the following relations are developed:

The existing design method assumes equal chord lengths, so:

When $L_{chord.(i-1)} = L_{chord.(i)} = L_{chord.(i+1)} = \cdots = L_{chord.(n)}$,

Then and for the same R_{min} values;

$$Dc._{(i-1)} = Dc._{(i)} = Dc._{(i+1)} = \cdots = Dc._{(n)}, \quad and; \tag{6.2}$$

$$Dc._{(i-1)} = allowable.Dc._{(i-1)} = Dc._{(i)} = allowable.Dc._{(i)} \ldots = Dc._{(n)}, \quad Then;$$

$$\Delta = Dc._{(i-1)} + Dc._{(i)} + Dc._{(i+1)} + \cdots + Dc._{(n)}$$

While,

$$allowable.Dc.(i) = Dc.(i-1)/2 + Dc.(i)/2$$

$$allowable.Dc.(i+1) = Dc.(i)/2 + Dc.(i+1)/2$$

And when $Dc._{(i-1)} = Dc._{(i)} = Dc._{(i+1)} = \cdots = Dc._{(n)},\ Then;$

$$allowable.Dc.(i) = Dc.(i)$$

$$Total\ length\ of\ transition = L_{chord.(i-1)} + L_{chord.(i)} + L_{chord.(i+1)} + \ldots + L_{chord.(n)}$$

Where, R_{min} is radius of the curve, D_c is the degree of the curve, Δ is the total deflection angle, and L_{chord} is the length of a chord.

It is important to mention that a smooth transition layout between two given tangents (T) is obtained by providing successive deflections (curvatures equal to D_c in value) to the successive chords (L_{chord}). Then the cumulative of (n) $D_c s$ and $L_{chord} s$ should satisfy a smooth transition. The total transition in curvature (cumulative of $D_c s$) would be equal to (Δ) (the total deflection angle required to make a total change in direction between the two tangents T1 and T2) and the total chord lengths would approximately be equal to the total length of the curve.

6.3 THE STATION POINT APPROACH AND HORIZONTAL CURVATURE

The concept of the station points in this study was integrated with the principles of horizontal curvature (mentioned above) to achieve smoother transitions in horizontal direction. It was suggested to use chord definition as the degree of the curve to orientate the required curvature without the need to use the traditional geometric design elements such as horizontal curves, tangents, and intersection points. The proposed concept tries to provide smooth transitions through the locations of the station points. The locations of the station points need to be orientated (positioned) in such a way so that the deflection between any three successive station points does not exceed the maximum allowable deflection defined by the distances between the points (chord lengths according to the traditional design method) and the defined radius (see equations 6.5.1 to 6.6.3). The ideas, the formulations, and the algorithms that are suggested for handling horizontal curvature as constraints for horizontal highway alignments are discussed in the following sections.

Different standards use different lengths for chord definition. The most common length is 100ft (30.5m). A shorter chord length definition will result in a smoother transition and safer driving conditions without compromising the design speed. Consequently, the chord length in this study, which was considered a key decision for the required curvature, was left as a user-defined parameter with a caveat to not exceed a maximum of 100ft (30.5m). The station point approach treats the distances between the points as the decision lengths (chord lengths) for the curvature configuration. Equation 6.1 above calculates the maximum allowable Dc when the full length of chord according to the definition (100ft or 30.5m) is used, while equation 6.3 defines maximum allowable D_c requirement as a function of variable chord lengths, provided that they are not greater than the maximum (100ft or 30.5m).

$$D_c = 2 * \sin^{-1}\left(\frac{L_{chord}/2}{R_{min}}\right) \tag{6.3}$$

The traditional design approach deals with intersection points (IPs), point of curvature (PC), point of tangency (PT), and tangent lengths to set out a highway horizontal curve on the ground and it mostly depends on equal chord distances and consequently equal deflections between the chords to form the transition. The station point approach, although not utilising geometric design elements like IPs and tangents, uses the intervals between the station points as chords and treats any successive three station points as a curve where the maximum allowable deflection must not be violated. As the configuration of the alignment itself depends on the station points' positions, the distances between the station points may well vary. These distances could have lengths greater than 30.5m or conversely have lengths equal to and even less than 30.5m. On the other hand, the smoothness of the alignment at these points depends on the deflection (curvature) that the position of any given three station points produces. When the deflection between these three points is zero; then the three points represent a straight section on the alignment. When the deflection angle has a value, it forms a curve and it is essential that the angle is not greater than the allowable maximum curvature value obtained from the corresponding chord lengths (the distances between the three station points) and the defined minimum radius (R_{min}). Figure 6.2 and equations 6.4.1 to 6.4.3 illustrate the idea of curvature consideration imposed on the proposed station point approach for smooth horizontal highway alignment configuration. Note that the decision distances (chord lengths) are typically variable when the concept of the station points is utilised.

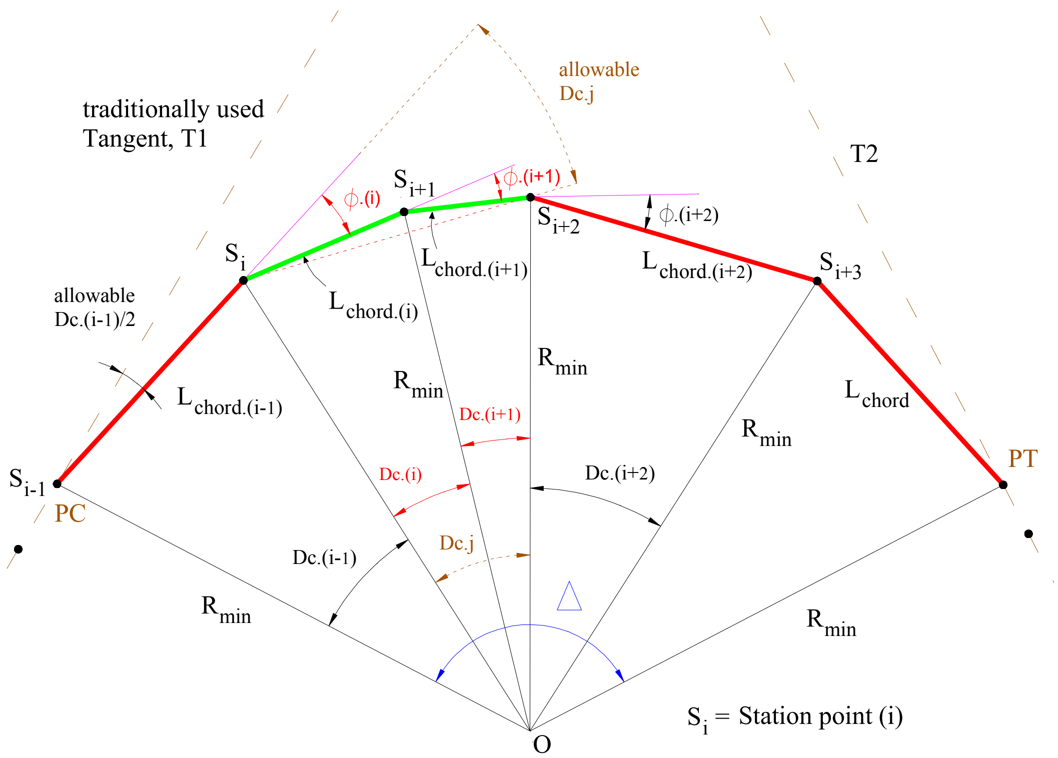


Figure 6.2: An illustration of curvature consideration with the concept of station point

When $L_{chord.(i-1)} \neq L_{chord.(i)} \neq L_{chord.(i+1)}$ ..., then; (6.4.1)

$D_{C.(i-1)} \neq D_{c.(i)} \neq D_{c.(i+1)}$ ..., then; (6.4.2)

$allowable.Dc_{(i-1)} \neq \emptyset.(i) \neq \emptyset.(i+1) \neq \emptyset.(i+2) \ldots$ (6.4.3)

With the station point approach, the distance inequality between the points might always be the case; therefore, individual allowable deflections at each station point need to be determined. For instance, taking the station points $S_i, S_{i+1}, S_{i+2},$ *and* S_{i+3} into consideration, the $D_c s$ are determined using equations 6.5.1, 6.5.2 and 6.5.3 (these are based on equation 6.3 above). Consequently, the maximum permissible curvatures that are required to make a smooth transition at the station points are determined using equations 6.6.1, 6.6.2, and 6.6.3. Note that these are examples for a few station points only and the same principle is used for the rest.

$$D_{c.(i-1)} = 2 * \sin^{-1}\left(\frac{L_{chord.(i-1)}/2}{R_{min}}\right) \quad (6.5.1)$$

$$D_{c.(i)} = 2 * \sin^{-1}\left(\frac{L_{chord.(i)}/2}{R_{min}}\right) \quad (6.5.2)$$

$$D_{c.(i+1)} = 2 * \sin^{-1}\left(\frac{L_{chord.(i+1)}/2}{R_{min}}\right) \quad (6.5.3)$$

$$\emptyset.(i) = D_{c.(i-1)}/2 + D_{c.(i)}/2 \quad (6.6.1)$$

$$\emptyset.(i+1) = D_{c.(i)}/2 + D_{c.(i+1)}/2 \quad (6.6.2)$$

$$\emptyset.(i+2) = D_{c.(i+1)}/2 + D_{c.(i+2)}/2 \quad (6.6.3)$$

The concept of station points calculates the above $(D_c)s'$ based on the actual lengths between the station points provided that the lengths are not exceeding 30.5m (100ft). When any of the distances between two station points exceeds 30.5m, the maximum chord length (30.5m), by definition, is used to determine the maximum allowable D_c (deflection or curvature) at the specified station point using equation 6.1. Any extra length beyond the 30.5m would be considered as a straight section with zero deflection with the preceding chord length and this keeps the curvature within the maximum allowable value. This notion assumes that the next station point locates at distance 30.5m where the required curvature is zero to the following station point.

The above equations tell that a deflection angle that is required to provide a curvature at a station point cannot be greater than the value determined by equation (6.1) for any chord length under any circumstances. Meanwhile, when the chord length is less than 30.5m, where equations 6.5.1 to 6.6.3 are used, the allowable maximum curvature would become smoother than that being found by equation (6.1).

The calculated maximum allowable curvatures at each station point will then be compared with the actual existing deflection resulting from the alignment configuration at the specified point.

6.3.1 ACTUAL (EXISTING) CURVATURE CALCULATION

The existing deflection angle value between the three successive station points is compared with the maximum allowable curvature. If the existing one is less than the allowable curvature then the alignment at this point is smoother than allowable smoothness and no action will be required. When the actual curvature is greater than the allowable value it indicates that the alignment violates the curvature requirements at the designated position and remedial action would be required.

In this study an algorithm was designed to go through the station points and calculate the existing curvature between any three successive points (between two chord distances). The following steps illustrate the procedures that were used to calculate the existing (actual) curvature at a station point:

Step 1: From Figure 6.3 consider the station points S_{i-1}, S_i, and S_{i+1}.

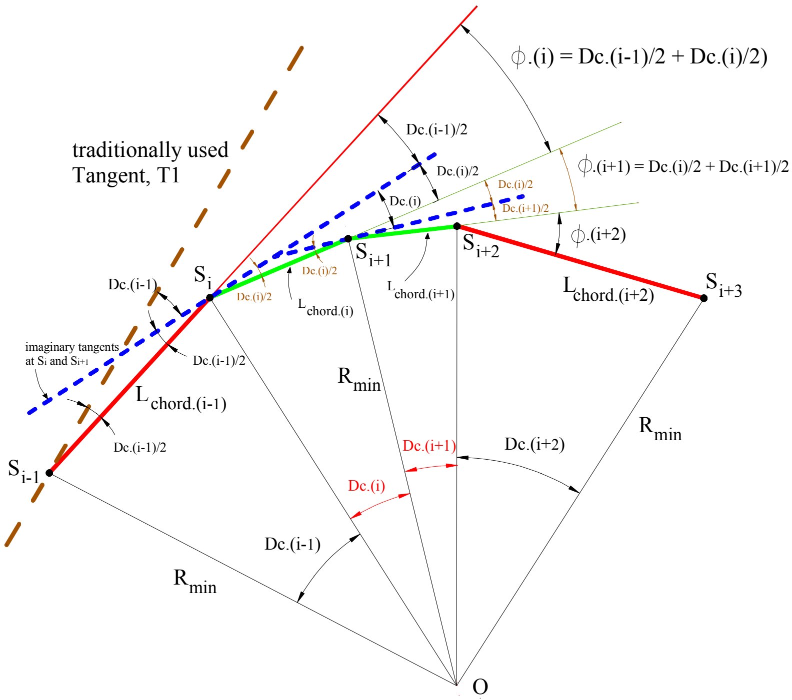


Figure 6.3: An illustration for existing curvature calculation

Step 2: Determine the length of vectors $(S_{i-1}S_i) = V_{i-1} = L_{chord.(i-1)}$ and$(S_iS_{i+1}) = V_i = L_{chord.(i)}$ (see Figure 6.4 for more detail).

$$a_x = X_{i-1} - X_i \tag{6.7.1}$$

$$a_y = Y_{i-1} - Y_i \tag{6.7.2}$$

$$b_x = X_i - X_{i+1} \tag{6.7.3}$$

$$b_y = Y_i - Y_{i+1} \tag{6.7.4}$$

$$V_{i-1} = \sqrt{(a_x{}^2 + a_y{}^2)} \tag{6.8.1}$$

$$V_i = \sqrt{(b_x{}^2 + b_y{}^2)} \tag{6.8.2}$$

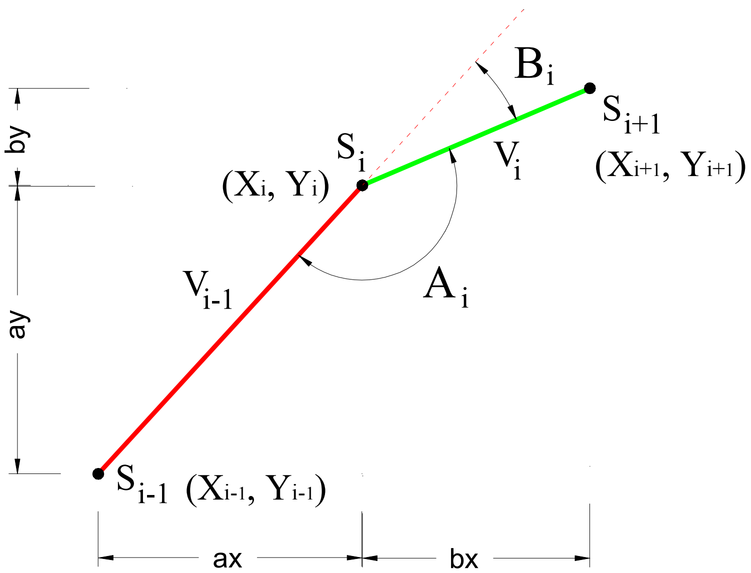


Figure 6.4: Illustration for actual deflection angle calculation

Step 3: Determine the actual deflection angle (Curvature) (B_i) between the two vectors V_{i-1} and V_i (Figure 6.4).

$$A_i = \cos^{-1}\left(\frac{(a_x * b_x + a_y * b_y)}{V_{i-1} * V_i}\right) * \left(\frac{180}{\pi}\right) \quad (6.9.1)$$

$$B_i = 180 - A \quad (6.9.2)$$

A_i is multiplied by 180/pi to convert the angle value from radian to degree. B_i is the actual deflection angle between the two vectors.

Then the calculated deflection angle B_i is compared with the maximum allowable curvature $\emptyset_i$. With this comparison the following cases would arise:

Case 1:

$$B_i \leq \emptyset_i$$

This means that the existing curvature is smoother than the maximum allowable curvature and therefore no action is required. Go one step forward and check curvature at the next station point considering S_i, $S_{i+1}, and\ S_{i+2}$.

Case 2:

$$B_i > \emptyset_i$$

This means that the existing curvature violates the maximum allowable curvature and therefore an action or actions is (are) required. In this study several actions were suggested to deal with a curvature that violates the maximum allowable value. These actions are discussed in the following sections.

Figure 6.5 illustrates the process of horizontal curvature consideration in a flowchart format.

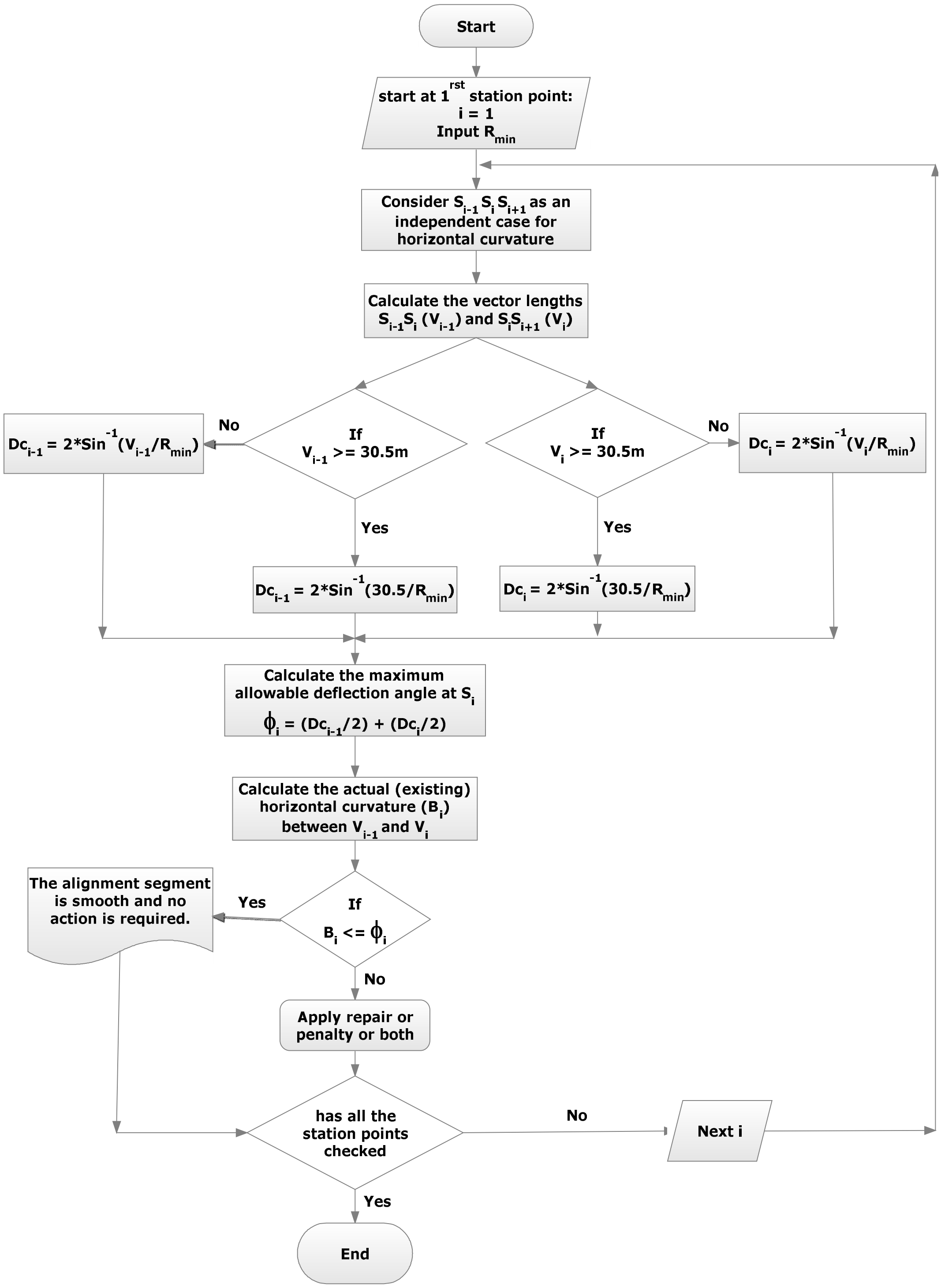


Figure 6.5: The process of horizontal curvature consideration with the concept of station points

6.3.1.1 Penalty Approach

Imposing a penalty on a candidate solution is one of the approaches adopted in this study. A soft penalty approach was suggested so that the model can distinguish between higher and lower levels of curvature violations. The total value of violated curvatures along the alignment at all the station points was calculated and multiplied by a penalty cost defined by the user, as follows:

$$when\ B_i > \emptyset_i\ then:$$

$$Total\ violated\ horizontal\ curvature\ value;\ VHCV_T = \sum_{i=1}^{N-2} (B_i - \emptyset_i) \qquad (6.10)$$

Where, N is the total number of station points. N-1 represents the end station and N-2 represents the point before the end station. This is done because the start station is represented with point number (0).

Then the cost that was imposed on the candidate solution was calculated as:

$$PC_{H\_Curvature} = VHCV_T * UHPC \qquad (6.11)$$

Where $PC_{H\_Curvature}$ is the total horizontal curvature penalty cost and UHPC is a user-defined Unit Horizontal Penalty Cost.

6.3.1.2 Repair Approach

This approach adopts a step by step repair procedure to gradually repair the positions of the station points where the curvature is violated. A step by step repair means that at each generation the violated curvatures along the alignment station points are reduced by an amount and the candidate solution is re-evaluated and reinserted into the population. In subsequent generations the same procedure is repeated so that at the end of the search a smooth horizontal alignment is achieved with no or only few curvature violations.

The algorithm that was formulated to perform repairing is illustrated in Figure 6.6 below:

For i = 1 to N-2 step 1

Calculate the maximum allowable curvature $\emptyset_i$ at station point S_i

Calculate B_i at station point S_i

If $B_i > \emptyset_i$

Repair curvature (reduce curvature by reallocating the position of station point S_i)

End if

Next i

Figure 6.6: An illustration of horizontal curvature repairing algorithm

To repair the violated curvature (B_i), the position of the station point, where the curvature is violated (S_i), is reallocated (S_{i_new}) so that the existing curvature is reduced (B_{i_new}) (see Figure 6.7 for detail). The steps below give the details.

Step 1: Find a new position S_i^- midway on a straight line between $S_{i+1}\ and\ S_{i-1}$ (Figure 6.7-a):

$$S_i^-(X_i^-) = X_{i-1} + \frac{(X_{i+1} - X_{i-1})}{2} \quad (6.12.1)$$

$$S_i^-(Y_i^-) = Y_{i-1} + \frac{(Y_{i+1} - Y_{i-1})}{2} \quad (6.12.2)$$

Step 2: Find a new position S_{i_new} midway on a straight line between S_i and S_i^- (Figure 6.7-b):

$$S_{i\_new}(X_{i\_new}) = X_i^- + \frac{(X_i - X_i^-)}{2} \qquad (6.13.1)$$

$$S_i^-(Y_{i\_new}) = Y_i^- + \frac{(Y_i - Y_i^-)}{2} \qquad (6.13.2)$$

Step 3: Calculate (B_{i_new}) as the new existing (actual) curvature at station point (S_{i_new}).

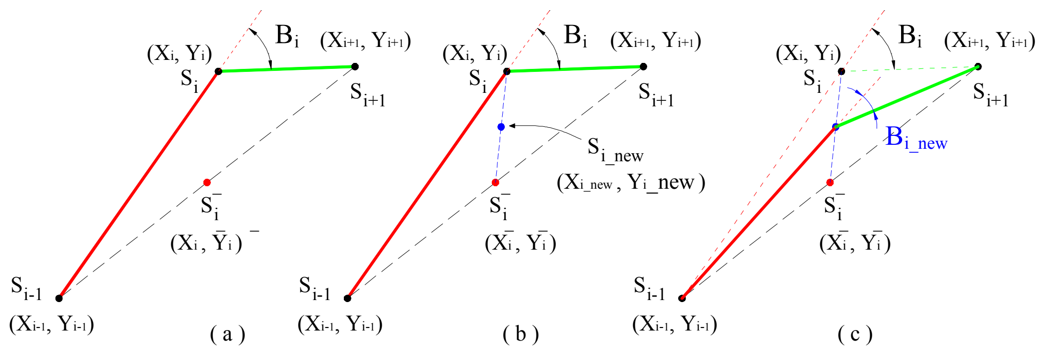


Figure 6.7: Illustration of the approach for horizontal curvature repairing

By using the above operations, smoother candidate solutions will be inserted into the next generation.

The algorithm in the proposed model is formulated in such a way that the actual and the maximum allowable curvatures at each station point of the optimum solution are presented (listed) on a windows form for user information, investigations, and decisions. For more details see Appendix 2 Figures App2.11 to App2.14.

It should be mentioned that the above two approaches (penalty and repair) were suited to three separate cases which were optional to the user, who was able to toggle between them. The different approaches were as follows.

- Penalty scenario: Adds the total penalty cost and the total horizontal curvature value (which were optional for the user) to the fitness function; or
- Repair scenario: Repairs the violated horizontal curvature values and adds the total horizontal curvature values (which were optional for the user) to the fitness function; or
- Repair and Penalty scenario: Repairs the violated horizontal curvatures, penalises the remaining horizontal curvature violations if they still exist after the repair process has been applied, and optionally adds the total horizontal curvature value to the fitness function.

6.4 THE VERTICAL CURVATURE AS A CONSTRAINT

Vertical alignments like horizontal alignments require the presence of smooth transitions between the different grade values. Traditionally, smooth vertical transition is accomplished using the geometric design elements of vertical highway alignment such as vertical intersection points, vertical tangents, gradients, algebraic difference between the different grades, vertical parabolic curves, K value, and design speed. These elements are used to transform the vertical gradients into elevations along the centreline of the horizontal alignment in such a way as to produce smooth and safe driving conditions. Special equations are available to deal with these elements and calculate their values (for more detail, see Chapter 2, Section 2.4.2).

With the concept of station points, different procedures were developed to produce smooth curvature between the different gradients. It was considered that the curvature at any point along the vertical alignment should satisfy the minimum requirements for safe driving. The details of the methods and the procedures are in the following sections.

6.5 THE STATION POINT APPROACH AND VERTICAL CURVATURE

In this study, the proposed station points approach tries to utilise a few of the traditional design elements and provide different formulations for smooth vertical transition. The design elements that required consideration where the concept of station points was concerned were design speed (V) and the K factor for both sag and crest vertical curves. The design speed is important to select a proper design value for K, and the K value is equally important to define the rate of vertical curvature. K is a minimum horizontal distance that is required to effect a maximum 1% change in the gradient along the vertical alignment curves (Figures 6.8). In other words, for the selected design speed, K is the minimum horizontal distance with which the change in curvature (the algebraic difference) between the successive grades along the vertical alignment should not exceed 1% $(e.g{:}\, |g_i - g_{i-1}| \leq 1\%)$. Traditionally, K is used to determine the minimum vertical curve length (L_v) between the two vertical gradients (G_1 and G_2), so that sufficient space is developed to distribute the total curvature (A) by amounts not greater than 1% per K horizontal distances. L_v is the result of multiplying A by K (see Equation 6.14).

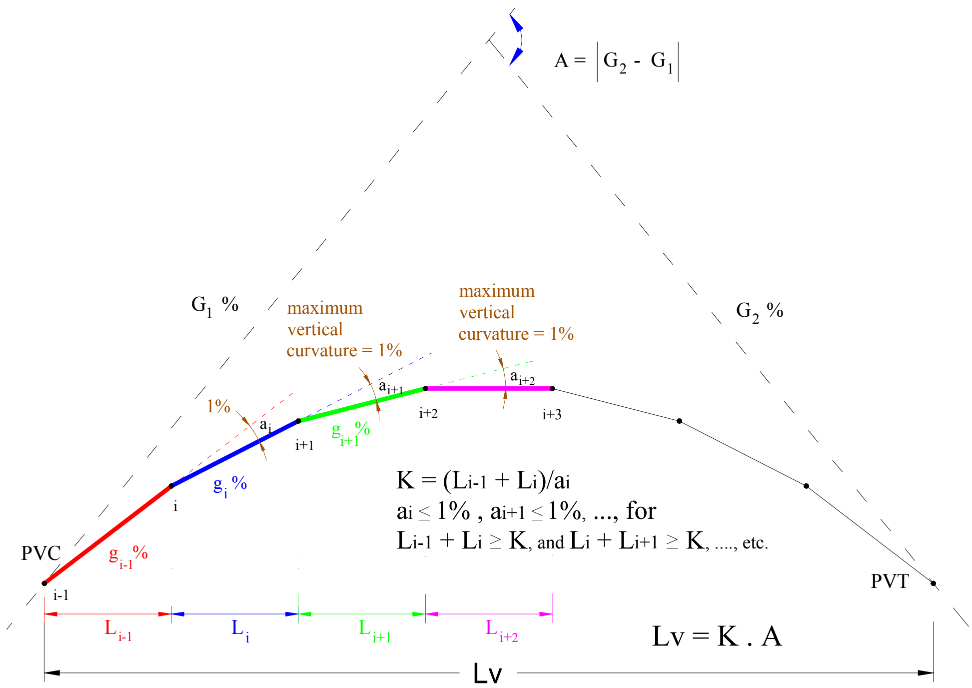


Figure 6.8: Illustration of K definition

The standard specification requirements for smooth and safe driving associate K values with the proposed design speed of the proposed alignment. To determine the limit of curvature or change in gradient between the successive grade lines, appropriate K values are selected according to the proposed highway design speed whereby the conditions like stopping sight distance (SSD) and SSD < Lv, which are considered essential to provide smooth and safe driving conditions, are satisfied (see Chapter 2, Section 2.4.2). The design values of K for sag and crest vertical curves are presented in Table 2.1 (Chapter 2) and in Appendix 1.

The following explains the basics of dealing with the K value so that it can be associated with the concept of station points for vertical curvature development.

Recall from the current traditional approach for vertical alignment design (Figure 6.8):

$$K = \frac{L_v}{A} \tag{6.14}$$

$$A = |G_2 - G_1| \tag{6.15}$$

Where L_v is the total vertical curve length between the two gradients $G_1 and\ G_2$ and A is the algebraic difference between the two grades.

K suggests that the gradient difference $(e.g.: a_i)$ between any two segment grades $(e.g.: g_{i-1}\%\ and\ g_i\%)$ along the vertical alignment curve must not be greater than 1%. Thus, for regular and equal curvature transitions, for example, between i-1, i, and i+1:

When $L_{i-1} = L_i = L_{i+1} = L = \cdots$ (6.16.1))

$$L_{i-1} + L_i = 2 * L \tag{6.16.2}$$

$$A = a_i = 1 \tag{6.16.3}$$

$$K = \frac{2 * L}{a_i} \quad (6.16.4)$$

$$K = 2 * L \quad (6.16.5)$$

$$L = \frac{K}{2} \quad (6.16.6)$$

Where; L is the half length of the curve between points i-1, i, and i+1 according to Equations 6.16.1 and 6.16.2.

Equation (6.16.6) implies that the curvature (gradient difference) between two segments (or gradients) must not exceed 1% when the segment lengths (L) are K/2 or more. In other words, for any gradient length equal to or greater than K/2 the maximum permissible change in gradient (curvature) is 1%. This notion or analysis was integrated with the proposed station point approach, which is now described in further detail.

From Figure 6.9, consider that S_{i-1}, S_i, S_{i+1}, S_{i+2}, ... are station points along the alignment and assume that each three successive points form an individual vertical curve case where the curvature should not exceed 1%. It should be mentioned that the individual vertical curve cases were overlapped such that the vertical curvature at every single station point was considered. The individual vertical curve cases were considered in the series of $(S_{i-1}\, S_i\, S_{i+1} = L_{v1})$, $(S_i\, S_{i+1}\, S_{i+2} = L_{v2})$, $(S_{i+1}\, S_{i+2}\, S_{i+3} = L_{v3})$, ... etc. Figure 6.9 illustrates the concept and also shows how the K value was represented. It should be noted that the illustration is for a case when the length of the curve (represented by three station points) is equal to or greater than K. Special cases are also considered when the lengths of vertical curves are less than K and when the distances between the station points are different. The details of the formulations for the different cases are presented in the following sections.

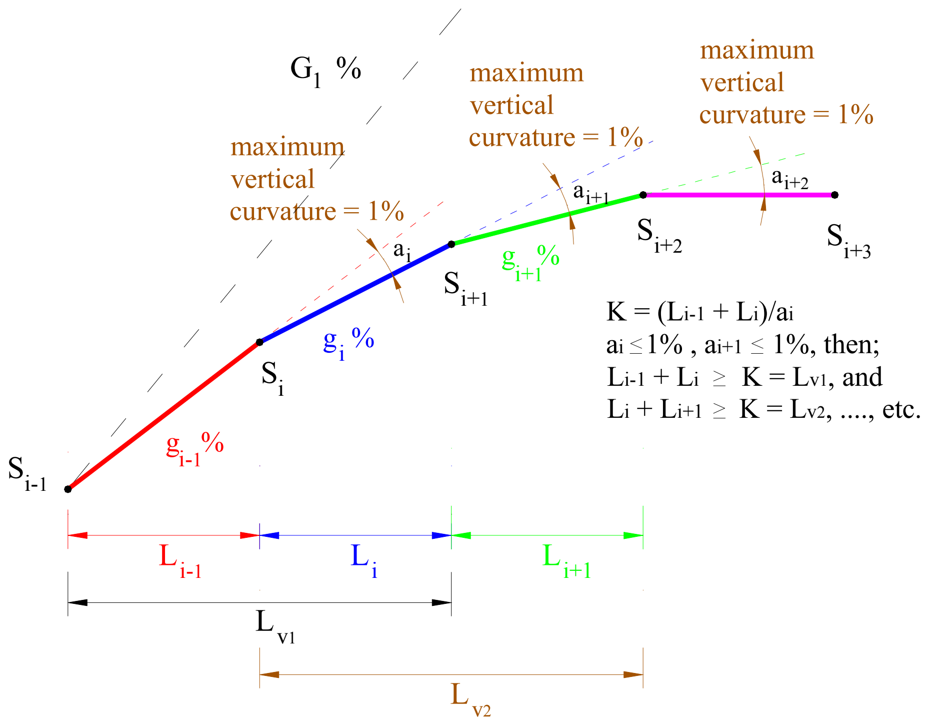


Figure 6.9: A typical illustration of vertical curvature with the concept of station point

6.5.1 ACTUAL VERTICAL CURVATURE CALCULATION AND THE MAXIMUM LIMIT

In this study the K definition principles were imposed on the proposed station point approach for vertical curvature satisfaction. The following steps describe the procedures for vertical curvature determination using the K factor.

Step 1

Consider three successive station points as an independent case for curvature consideration.

From Figure 6.9, let $S_{i-1}, S_i, and\ S_{i+1}$ be the three successive station points.

Step 2

Calculate the horizontal projection distances $(\mathrm{L_{i-1}} \text{ and } \mathrm{L_i})$ of the vectors $\mathrm{S_{i-1}S_i}$ and $\mathrm{S_i}\ \mathrm{S_{i+1}}$.

Step 3

Calculate the grade values $g_{i-1}\%\ and\ \mathrm{g_i}\%$ of line segments $\mathrm{S_{i-1}S_i}$ and $\mathrm{S_i}\ \mathrm{S_{i+1}}$.

Step 4

Select the least horizontal length between L_{i-1} and $\mathrm{L_i}$. This is suggested to overcome the worst case that could occur with various L lengths. Refer to equation (6.16.6) and step 6 for justification.

$$L_{min} = \min\ [L_{i-1}, L_i] \tag{6.17}$$

Step 5

Find out whether the line segments $S_{i-1}, S_i, and\ S_{i+1}$ act as a sag vertical curve or crest vertical curve:

If $(g_i - g_{i-1}) > 0$ then the curve or alignment segment is considered as Sag.

If $(g_i - g_{i-1}) < 0$ then the curve or alignment segment is considered as Crest.

This is required so that a proper K value is selected.

Step 6

Find whether L_{min} is greater or less than K/2.

$$L_{min} \geq \frac{K}{2} \ or \ L_{min} < \frac{K}{2} \tag{6.18.1}$$

If $L_{min} \geq \frac{K}{2}$ then go to step 7, if $L_{min} < \frac{K}{2}$ then go to step 8.

Step 7

When $L_{min} \geq \frac{K}{2}$

Calculate the actual (existing) algebraic gradient difference a_i between g_{i-1} and $\mathrm{g_i}$ and compare it with the maximum allowable gradient difference ($a_{\max}$) (see Figure 6.10):

$$a_{max} = 1\% \tag{6.19}$$

And the actual (existing) gradient difference is:

$$a_i = |g_{i-1} - g_i| \tag{6.20}$$

a_i $must\ be$ $\leq a_{max}(1\%)$ to provide a smooth and safe driving condition. Otherwise, apply penalty (see Section 6.5.1.1) or repair (see Section 6.5.1.2) or both.

Step 8

When $L_{min} < \frac{K}{2}$, a smoother curvature than the maximum ($a_{max} = 1\%$) would be required.

From Equation (6.16.4) above, the a_{max} is calculated as below (for more clarification, see Figure 6.10):

$$a_{max} = \frac{2 * L_{min}}{K} \tag{6.21}$$

And the actual (existing) gradient difference is:

$$a_i = |g_{i-1} - g_i| \tag{6.22}$$

a_i *must be* $\leq \left[a_{max} = \left(\frac{2 * L_{min}}{K}\right)\right]$ to provide a smooth and safe driving condition. Otherwise, repair (see Section 6.5.1.2) or apply penalty (see Section 6.5.1.1) or both.

This explains that, when the gradient length projection (L_{min}) is shorter than the minimum allowable length $(\frac{K}{2})$, which is defined as a minimum distance required to effect a maximum 1% gradient change, a smoother transition or change in gradient would be required to maintain the same driving condition. Equation 6.21 above produces a transition or curvature that is smaller (smoother) than 1%.

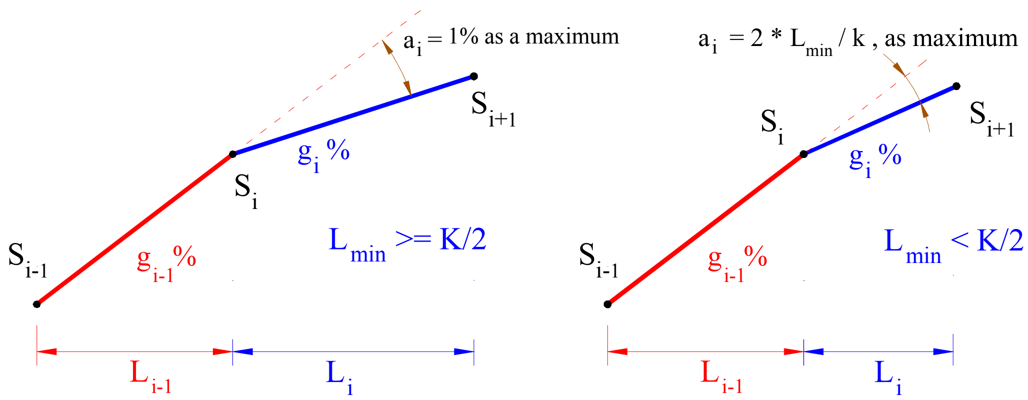


Figure 6.10: Vertical curvature configuration with different alignment segment length values

When the maximum allowable gradient difference $(a_{\max})$ and the actual curvature (a_{i}) are compared, one of the following cases would be inevitable.

Case 1: When

$$a_i \leq a_{max}$$

This implies that the curvature is smoother than the maximum limit and therefore no action is required. Continue checking the next station point vertical curvature until all the points have been checked.

Case 2: When

$$a_i > a_{max}$$

Having a steeper curvature than the maximum allowable limit necessitates taking an action (penalty and/or repair) to produce an alignment that meets the standard requirement for safe and smooth driving conditions.

Figure 6.11 shows the process of vertical curvature consideration in a flowchart format.

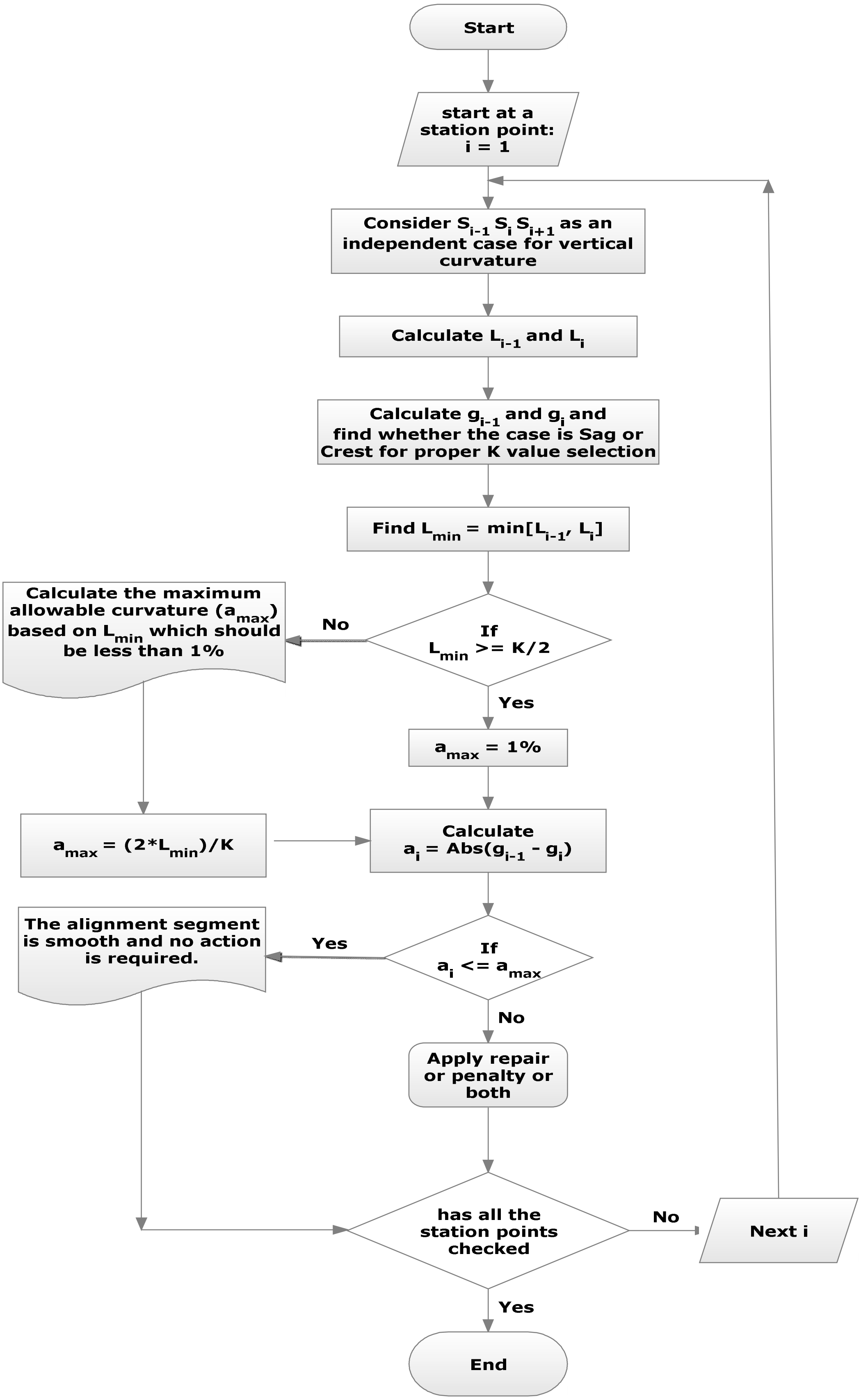


Figure 6.11: The process of vertical curvature consideration with the concept of station point

The following actions were suggested to deal with the cases where the vertical curvature violates the maximum allowable limits.

6.5.1.1 Penalty Approach

A similar approach to the horizontal curvature violation was adopted. A soft penalty was imposed on alignment solutions where vertical curvature violates the allowable gradient changes at their station points. The vertical penalty cost was calculated using the formulations below:

$$when\ a_i > a_{max}\ then:$$

$$Total\ violated\ vertical\ curvature\ value;\ VVCV_T = \sum_{i=1}^{N-2}(a_i - a_{max}) \qquad (6.23.1)$$

Where, N is the total number of station points. N-1 represents the end station and N-2 represents the point before the end station. This was done because the start station is represented by point number (0).

Then the cost that is imposed on the candidate solution would be:

$$PC_{V\_Curvature} = VVCV_T * UVPC \qquad (6.23.2)$$

Where $PC_{V\_Curvature}$ is the total vertical curvature penalty cost and UVPC is a user-defined Unit Vertical Penalty Cost.

6.5.1.2 Repair Approach

This approach adopts a step by step repair procedure to gradually repair the positions of the station points where the curvature is violated. A step by step repair means that at each generation the violated curvatures along the alignment station points are reduced by an amount and the candidate solution is re-evaluated and reinserted into the population. In the next generation, the same procedure is repeated so that at the end of the search a smooth vertical alignment is achieved with little or no curvature violations.

In this study a specific procedure was developed to repair the position of the station points where the vertical curvature violates the maximum allowable limit defined by a_{max}. Repairing the positions of the station points was limited to the elevation (Z coordinate values) of the station points, regardless of the station points' horizontal distance values (the X and Y coordinates of the station point). It should be mentioned that, as with horizontal curvature repairing, the repairs for deviations in vertical curvature were performed in steps, in that the selected (violated) station point was moved one step leaving the remaining violation (if any) for the subsequent generations. The following algorithm (Figure 6.12) describes the applied procedure that was used to improve the vertical curvature.

```
For i = 1 to N-2   step 1
Calculate the maximum allowable gradient difference (a_max) at station point S_i
Calculate the actual (existing) vertical curvature a_i at station point S_i
If a_i > a_max
Determine whether the alignment segment has Sag or Crest properties
Repair curvature according to the segment type "Sag or Crest" (reduce curvature
by adjusting the elevation of station point S_i)
End if
Next i
```

Figure 6. 12: An illustration of vertical curvature repairing algorithm

Considering a station point (see Figure 6.13), let it be (S_i), when the existing vertical curvature is greater than the maximum allowable $(a_i > a_{max})$, the elevation (Z_i) needs manipulation (Z_{i_new}) so that the existing curvature (a_i) is reduced (a_{i_new}). The steps below explain the detail.

Step 1: Consider the horizontal distance (d_i) (d_i is the station distance of station point S_i from the start point along the alignment) and elevation (Z_i) of station point (S_i) where the curvature (a_i) violates the maximum limit (a_{max}). Find the elevation (Z_i^-) at distance (d_i) on the straight line between S_{i+1} *and* S_{i-1} (Figure 6.13-b).

If $g_i - g_{i-1} > 0$ then the segment has Sag properties and $K = K_{sag}$

If $g_i - g_{i-1} < 0$ then the segment has Crest properties and $K = K_{crest}$

$$when\ L_{min} \geq \frac{K}{2}$$

$$a_{max} = 1\%\ and\ a_i = |g_i - g_{i-1}|$$

$$and\ when\ L_{min} < \frac{K}{2}$$

$$a_{max} = \frac{2 * L_{min}}{K} \quad and \quad a_i = |g_i - g_{i-1}|$$

$$then\ when\ a_i > a_{max}$$

$$Z_i^- = Z_{i-1} + \left(\frac{(Z_{i+1} - Z_{i-1})}{(L_{i-1} + L_i)}\right) * L_{i-1} \tag{6.24}$$

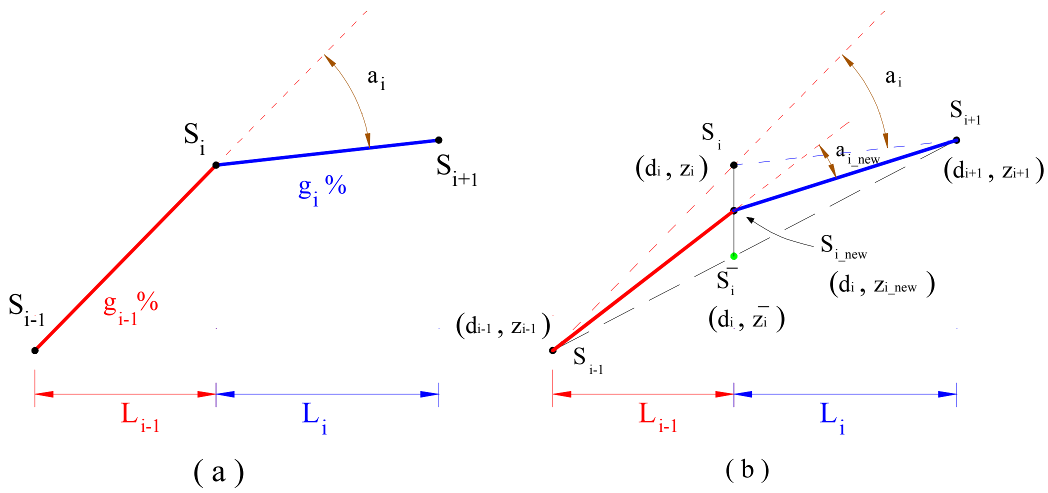


Figure 6.13: Illustration of repairing vertical curvature

Step 2: Find a new elevation Z_{i_new} midway on a straight line between S_i and S_i^-:

$$Z_{i\_new} = \frac{(Z_i + Z_i^-)}{2} \quad (6.25)$$

Step 3: Calculate (a_{i_new}) as the current vertical curvature at station point (S_{i_new}).

The application of the above repair procedures will result in a smoother vertical alignment with no or few vertical curvature violations. This process will continue with the candidate solutions until the search process ends.

The user of this model can check the vertical curvatures through a Windows form report, which is designed specifically for this purpose. The report provides the station point number, chord length, maximum allowable vertical curvature, and the actual (existing) vertical curvature. Moreover, the report shows whether the existing curvature violates the maximum permissible value or not.

In this study, the two approaches described above are such that they can be applied to suit three different scenarios depending on user preference, i.e. the selection of each scenario is made optional by the model user. These scenarios are:

- Penalty scenario: Adds the total penalty cost and the total vertical curvature value (which is optional for the user) to the fitness function; or
- Repair scenario: Repairs the violated vertical curvature values and optionally adds the total curvature value to the fitness function; or
- Repair and Penalty scenario: Repairs the violated vertical curvatures, penalises the remaining violated curvatures after the repairs have been applied, and optionally adds the total vertical curvature value to the fitness function.

6.6 MAXIMUM VERTICAL GRADIENT AS A CONSTRAINT

One of the considerations of vertical alignment is the gradient. A maximum allowable grade value, which is defined by standard specifications, is suggested by the user or planner and the gradient of any alignment segment between the successive station points should not exceed this value. The planner or the user needs to determine the maximum allowable grade value according to the type of road, topography (plain, rolling, and mountainous), weather condition where the road is being built, and the design speed. Keeping the vertical alignment gradients within the defined one is important to maintain the design speed, especially for heavy vehicles. Furthermore, a low gradient reduces the risk of traffic accidents.

In this study the effect of high grade values was considered by adopting a soft penalty approach, as in the following steps.

Step 1: Calculate the actual slope (grade) $g_{i\_actual}\%$ of line segment between station points (i) and (i+1)

$$g_{i\_actual} = \frac{(Z_{i+1} - Z_i)}{(d_{i+1} - d_i)} * 100 \qquad (6.26.1)$$

Where i = 0, 1, 2, ..., N-2

Step 2: Compare the absolute value of the actual grade $|g_{i\_actual}\%|$ with the maximum allowable grade $G_{max}\%$

If $|g_{i\_actual}\%| > G_{max}\%$

Then apply a soft penalty. The grade penalty function may take the form of the following equation:

$$PC_{Grade} = \sum_{i=0}^{i=N-2} \left(|g_{i\_actual}| - G_{max}\right) * UDGC \qquad (6.26.2)$$

Where PC_{Grade} is the total penalty cost that is imposed on the alignment due to vertical grade violations, g_{i_actual} is the actual gradient of the alignment segment between station points (i) and (i+1), and UDGC is a User Defined Grade unit Cost.

6.7 BALANCED CUT AND FILL EARTHWORK AS A CONSTRAINT

The costs of earthwork for cut and fill, as one of the fitness function components, could create significant differences between the candidate solutions. The alignment location, configuration, and total cost may be affected by the earthwork costs. The independent costs of cut and fill may also produce different answers to the alignment location and configuration. Soil conditions, availability of materials, and hauling distances (hauling costs) are factors that affect the total earthwork costs. However, from the economical viewpoint, the highway planners still prefer a solution in which the quantities of cut and fill are balanced (refer to Chapter 2, Section 2.10.1.3).

The model of this study was designed in such a way that preferences can be given to cut or fill or a balance between cut and fill. The model user can also give no preferences to either item and allow the process to make a free and a relaxed search.

The model application for giving preferences to cut or fill costs are straightforward, in which the cut or fill costs can be given a higher weighting factor. For instance, if it is desired to build a highway in an area where more cut than fill is demanded (e.g. where the surface soil needs to be removed due to poor condition or when the fill material costs are too high), a high weighting factor will be given to the fill costs (a_{fill} in equation 6.27 below) so that the search biases towards cuts and tries to avoid making fill sections (this was tested and presented in the experiment sections).

$$C_{Earthwork} = a_{Cut} * C_{Cut} + a_{Fill} * C_{Fill} \qquad (6.27)$$

Where $C_{Earthwork}$ is the total earthwork cost of the alignment, C_{Cut} is the total cut cost, C_{Fill} is the total fill cost, a_{Cut} is the cut cost weighting factor, and a_{Fill} is the fill cost weighting factor. Note that this equation is modified over equation 5.5.7 presented in Section 5.3.2 of Chapter 5. Equation 5.5.7 was used without giving weight to cut and/or fill costs and they were treated as total earthwork costs only.

Moreover, in order to use the model so that the earthwork quantities of cut and fill are balanced, a soft penalty equation was formulated in which the differences of both quantities were calculated and multiplied by a user-defined penalty cost, as shown in equation 6.28 below:

$$PC_{Earthwork} = |Q_{Cut} - Q_{Fill}| * UDPC_{P\_Earthwork} \quad (6.28)$$

Where $PC_{Earthwork}$ is the total penalty cost for the difference between cut and fill quantities (Q_{cut} and Q_{fill}), and $UDPC_{P\_Earthwork}$ is a User Defined Penalty Cost.

In this study the selection of these earthwork-related items was made optional by the user of the model. Different types of planning and design require different requirements to be satisfied. Several scenarios were considered, and through each the sensitivity of the model upon different strategies was tested.

6.8 THE FITNESS FUNCTION

The fitness function that has been developed for a 3D highway alignment model (Chapter 5) was reformulated to include the penalty costs of horizontal curvature, vertical curvature, grade, and earthwork balance for cut and fill, as follows:

$$Total\ Fitness\ Function: Minimise\ the\ total\ cost$$

$$= a_1.C_{Length} + a_2.C_{Location} + a_3.C_{Earthwork} + a_4 * PC_{H\_Curvature} + a_5 * PC_{V\_Curvature} + a_6 * PC_{Grade} + a_7 * PC_{Earthwork} \quad (6.29)$$

6.9 THE EXPERIMENTS AND THE RESULTS

Several test scenarios were prepared to test the efficiency of the model for handling different constraint components. For this purpose, it was suggested to use the world W_9 (as in the 3D model in Chapter 5) to allow comparisons between the results that are obtained with the constraint components and with those obtained with no constraint consideration (see the 3D model in Chapter 5).

It was aimed by the tests of this section to find an optimum 3D highway alignment that conforms to the horizontal curvature, vertical curvature, and maximum gradient requirements. Furthermore, the model was tested against the requirement for a highway alignment solution that balances out the cut and fill quantities.

The design and model parameters found for 3D highway alignment (as in the model in Chapter 5) were used as a base for the tests of this section (Table 6.1). The tests described in this chapter complement the tests made in the previous chapters and attempts to tune the parameters that yield a smooth and realistic 3D alignment according to horizontal curvature, vertical curvature, and gradient. Moreover, the requirement for a balanced cut and fill was also tested.

Table 6.1: The base parameters for 3D highway alignment solution*

P_{size}	N	CrO_Type	MM	SS	Z consideration
5000	60	RMPCrO-8P	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	RndSS	Z Initialisation within feasible limit defined by G_{max}%

(* Note: These parameters are taken from the result conclusions of Chapter 5)

The test scenarios of this section were made different through the total generation number (G) and the starting generation number for the application of penalty and repair procedures for horizontal and vertical curvature handling. The highway alignment design parameters and the parameters of the test scenarios are listed in Tables 6.2 and 6.3.

Table 6.2: The design parameter values of the different test scenarios

Design speed	60 mph
K value for Sag vertical curves	40.8 m
K value for Crest vertical curves	45.3 m
Chord length according to chord definition for degree of the curve	30 m
Minimum radius for horizontal curvature	150 m
UHPC	1 unit cost
UVPC	1 unit cost
UDGC	1 unit cost
$UDPC_{p\_Earthwork}$	1 unit cost

Table 6.3: The parameters of the different test scenarios

Test scenarios	G	Start generation number (g) for applying repair and penalise techniques for Horizontal and Vertical Curvatures				
S1	3000	0	25	50	75	100
S2	1000	0	25	50	75	100
S3	500	0	25	50	75	100

As mentioned earlier, due to the probabilistic nature of the GA-based model and to understand the behaviour of the solution algorithm in a broader vision, the program was run 10 times per test scenario per parameter (start generation number for applying repair and penalise was considered as a parameter). For instance, the total number of test runs for the test scenarios in Table 6.3 was 150. In other words, the tests were divided into three groups (scenarios) each with five different parameters (g).

The aim behind using a different generation number was due to the way of considering the boundary area limits for mutation. The applied mutation method with UEA(1/1) reduces the range of mutation by time with which each different generation number produces different mutation area limits, as shown in Figure 6.14. By this approach, the behaviour of the model will be shown under a different area consideration for mutation. The selected parameters (g as a start point for applying repair and penalise) will also investigate the time required to start applying repair and penalise (R&P) techniques for horizontal and vertical curvatures. Figures 6.15 to 6.20 show the results using different presentation formats.

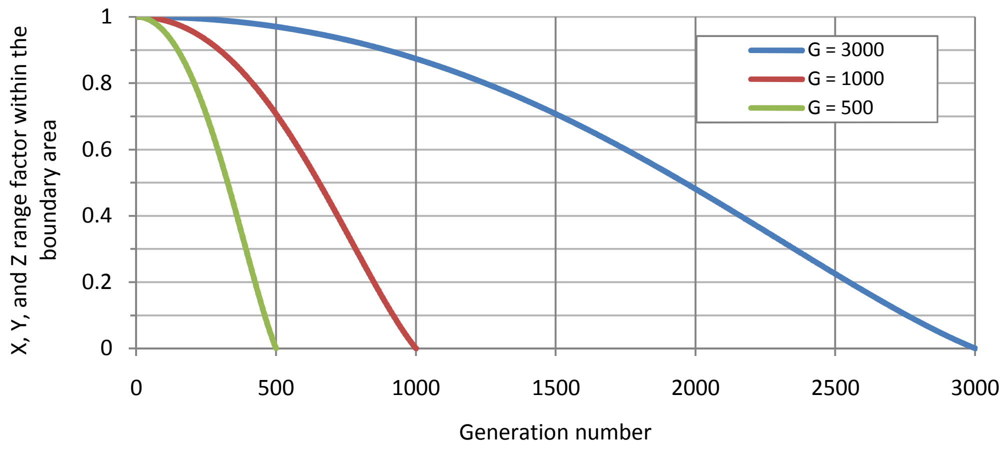


Figure 6.14: Mutation range factor for area consideration using UEA(1/1)

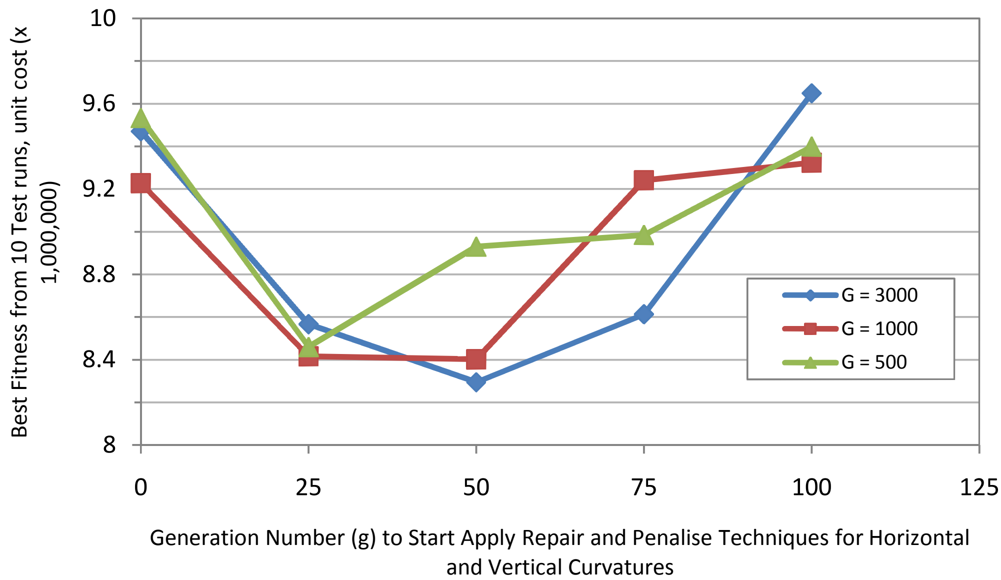


Figure 6.15: The effect of start generation number to apply R&P techniques

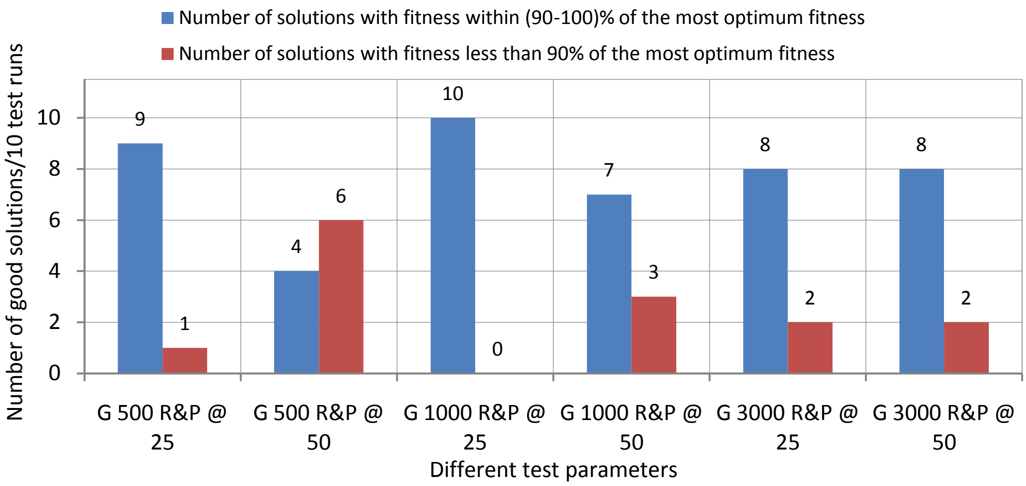


Figure 6.16: The effect of different parameters on the stableness of the solutions obtained from 10 test runs

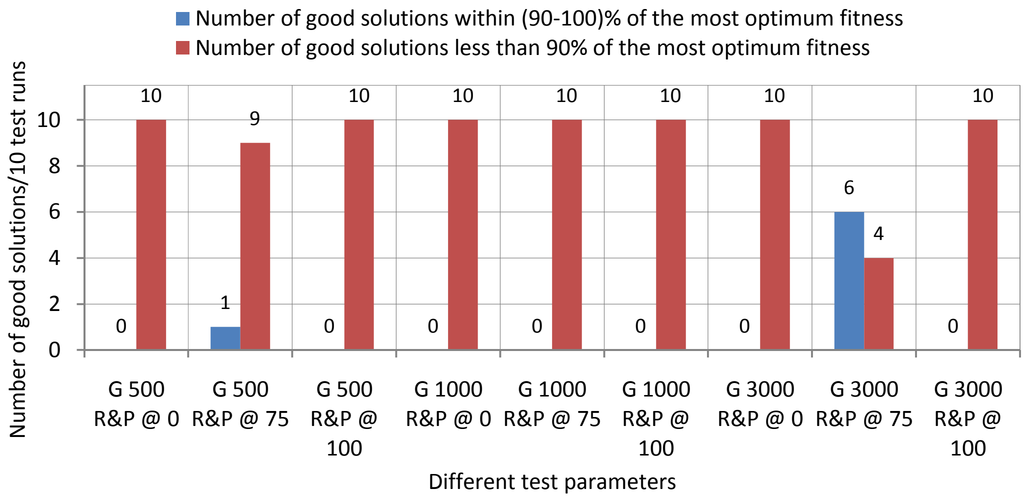


Figure 6.17: The effect of different parameters on the stableness of the solutions obtained from 10 test runs

Figures 6.15 to 6.17 show that the best solution can be obtained with parameters R&P@25 and R&P@50. Moreover, the solutions with G 1000 and 3000 were fitter than those obtained with G 500. Figure 6.18 shows the best fitness graphs for the parameters that produce the most stable results. Meanwhile, Figure 6.19 reveals the differences between the minimum and maximum fitness values for the solutions that use these parameters.

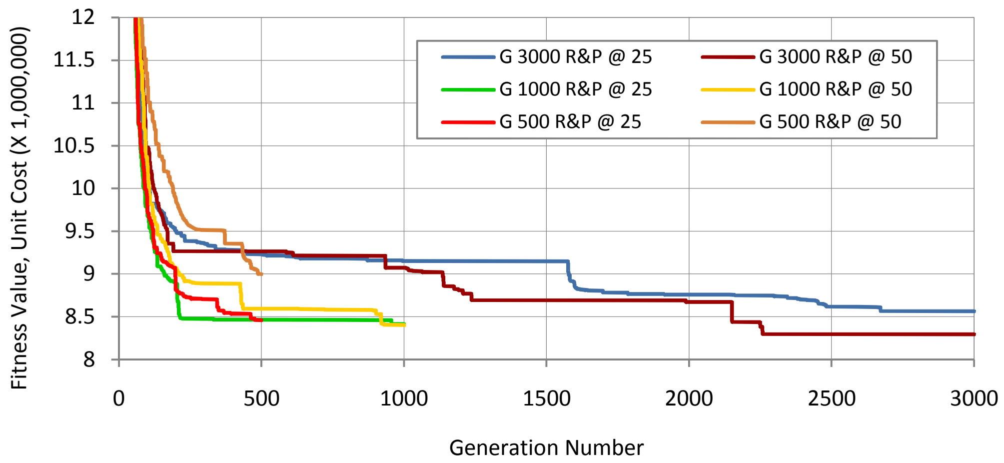


Figure 6.18: The best fitness values

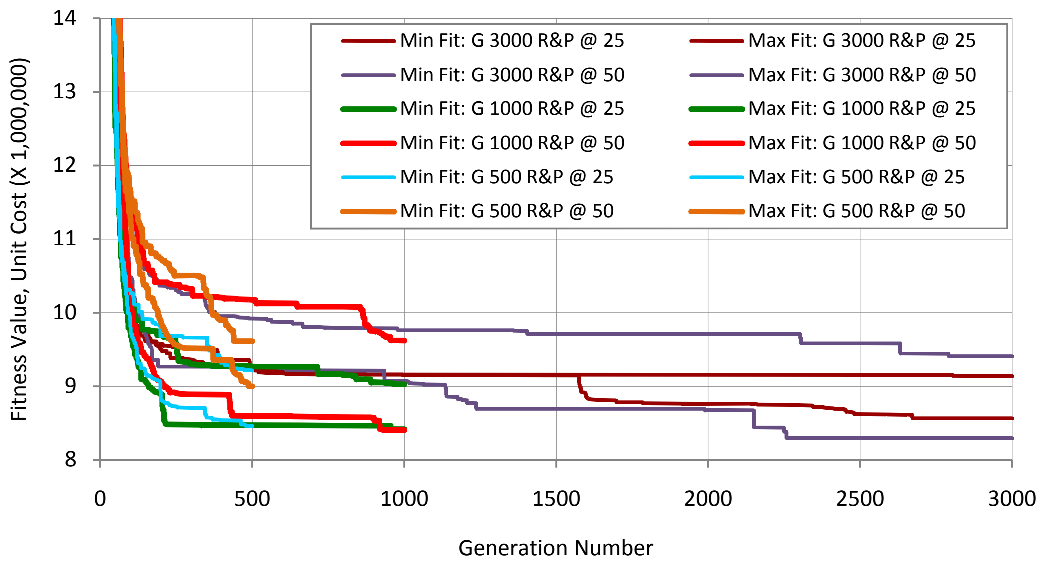


Figure 6.19: Minimum and maximum fitness graph

As the aim was to produce a smooth and realistic 3D highway alignment in conjunction with the best fitness and the number of good solution results, the other alignment characteristics (e.g. horizontal and vertical curvature measures) therefore needed to be observed and analysed. Horizontal and vertical alignment-related characteristics that were considered for the alignment evaluation are listed and abbreviated in Table 6.4. The values of these elements for the best solutions are shown in Table 6.5.

This analysis was done to recognise the most efficient parameter set that results in producing better solutions.

Table 6.4: Description of the horizontal and vertical alignment characteristics

Horizontal Alignment	
No VHC	Number of station points where horizontal curvatures are violated (VHC: Violated Horizontal Curvature) (No).
THC	Total Horizontal Curvature Measurement along the horizontal alignment (degree).
HCI	Horizontal Curvature Index: Is the ratio of the total horizontal curvature to the alignment length ($HCI = THC/L$).
Max_C_P	Maximum curvature value and its position. The position is the station point location where the maximum curvature occurs.
AHC	Allowable Horizontal Curvature (degree).
TAOVHC	Total Amount Of Violated Horizontal Curvature (degree).
Vertical Alignment	
No VVC	Number of station points where vertical curvatures are violated (VVC: Violated Vertical Curvature) (No).
TVC	Total Vertical Curvature Measurement along the vertical alignment (degree).
VCI	Vertical Curvature Index: Is the ratio of the total vertical curvature to the alignment length ($VCI = TVC/L$).
DiffCutFill	The amount of difference between Cut and Fill quantities (m^3).
Exist G_{max}	Maximum existing grade value (%).
A-G_{max}	Maximum allowable grade value (%).
TAOVVC	Total Amount Of Violated Vertical Curvature (degree).

Table 6.5: The results for horizontal and vertical alignment characteristics

	G 500 R&P@25	G 500 R&P@50	G 1000 R&P@25	G 1000 R&P@50	G 3000 R&P@25	G 3000 R&P@50
	Horizontal Alignment Records					
No VHC	4	2	5	8	4	2
THC	467.7	426.35	405.4	414.75	440.85	462.1
HCI	0.0321	0.0294	0.0283	0.0289	0.0306	0.032
Max_C_P	27.3@8	28.9@10	18 @ 6	25.8@ 9	29.7@10	51.1@10
AHC	11.47	11.47	11.47	11.47	11.47	11.47
TAOVHC	41.47	34.48	20.47	44.68	28.6	71.84
	Vertical Alignment Records					
No VVC	4	0	6	9	9	10
TVC	36.47	32.02	38.79	41.92	39.1	40.8
VCI	0.0025	0.0022	0.0027	0.0029	0.0027	0.0028
DiffCutFill	9,480	38,953	20,110	29,989	2,505	11,454
Exist G_{max}	4.35%	3.9%	4.5%	6.05%	4.17%	5.11%
A-G_{max}	8%	8%	8%	8%	8%	8%
TAOVVC	5.28	0	5.42	11.47	3.88	9.16

Comparing the overall performances gives almost the same preferences to the selected parameters. Figures 6.20–6.22 graphically show the best results produced from G1000 R&P@25. The alignment's fitness of the presented solution is 8,415,960 unit cost and it is 98.53% optimum as the most optimum solution of fitness (8,294,256). This solution is chosen based on the characteristics shown in Table 6.5.

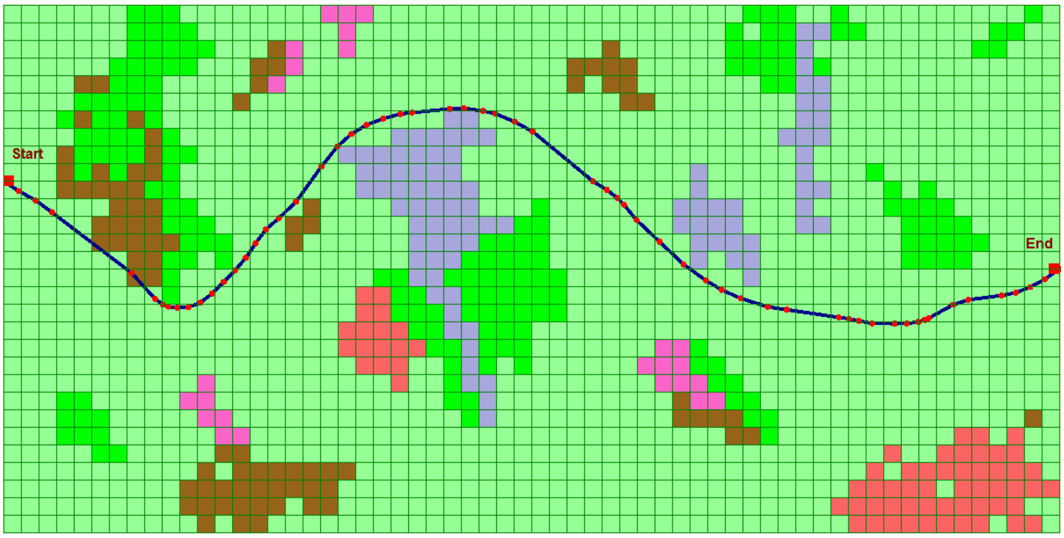


Figure 6.20: The optimum horizontal highway alignment test result on a land use grid map

(G1000 R&P @25)

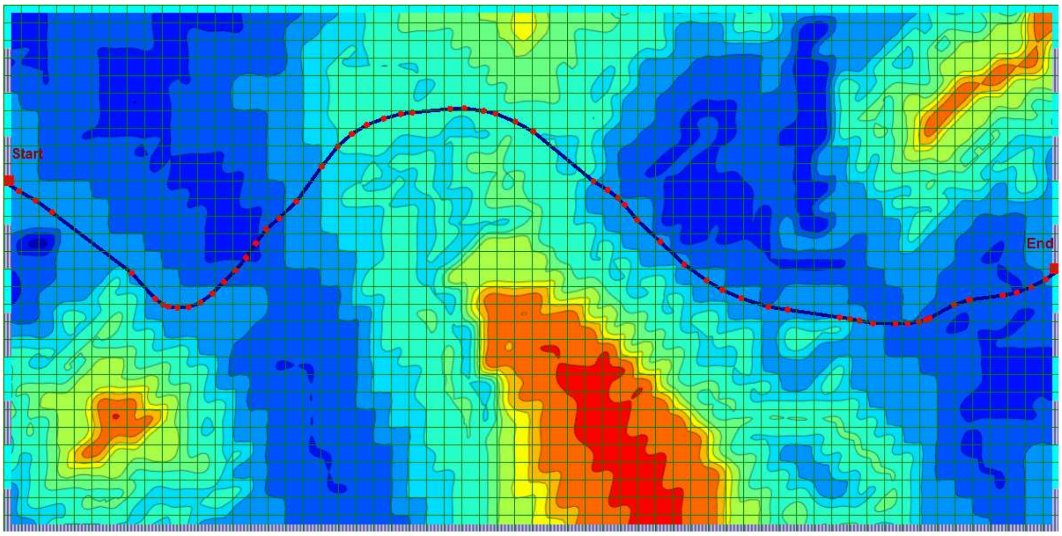


Figure 6.21: The optimum horizontal highway alignment test result on a contour map

(G1000 R&P @25)

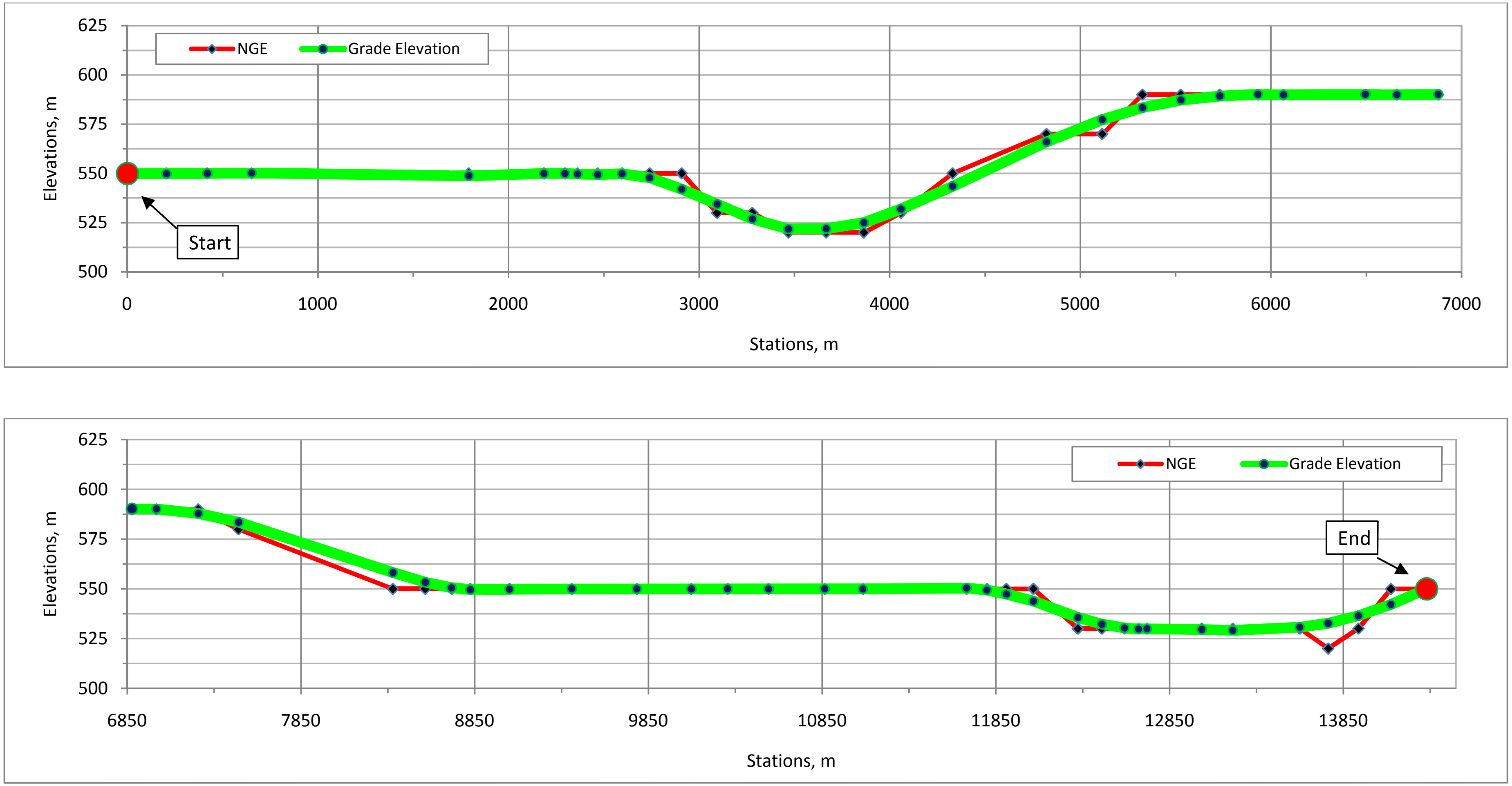


Figure 6.22: The profile of the vertical highway alignment

Table 6.6 below compares the optimum (or near optimum) solutions resulting from the tests considered with and without the constraint components' handling. The table shows the differences between the performances of the two models for 3D highway alignment optimisation. For comparison refer also to Figures 5.37– 5.39 in Chapter 5.

Table 6.6: Comparison of the test results with and without constraint handling

	S7 from Chapter 5 without constraints for curvature, gradient, and earthwork balancing	Parameter set G1000 R&P@25 of this chapter with curvature, gradient, and earthwork balancing
Fitness Value	6,771,815 unit cost	8,415,960 unit cost
Cut volume	23,192 m^3	134,370 m^3
Fill volume	44,431 m^3	114,260 m^3
Total Cut – Fill volume	67,623 m^3	248,630 m^3
Difference in Cut and Fill	21,239 m^3	20,110 m^3
% of Cut and Fill difference to the total earthwork amount	31.4%	8.08%
Exist G_{max}	-118.29%	4.58%
THC	802.642 degree	405.4 degree
HCI	0.0605 degree/unit length	0.0283 degree/unit length
TVC	228.77 degree	38.79 degree
VCI	0.0173 degree/unit length	0.0027 degree/unit length

The results explain that satisfying vertical gradient requirement caused more cut and fill quantities due to the maximum allowable grade limitations. However, the percentage of cut and fill differences to the total earthwork amount, due to the requirement for balancing cut and fill quantities, was much smaller than if this balancing was not considered. The comparison shows that the horizontal and vertical alignment can be made smoother with the techniques adopted in the study.

So, the conclusions were drawn to be:

- At least 1,000 generation numbers are needed to optimise 3D highway alignment when constraints are handled.
- Based on the results from the test scenarios, the repair and penalisation algorithms need to be applied between generations 25 and 50.

6.10 SENSITIVITY AND VERIFICATION TESTS

6.10.1 SENSITIVITY TESTS

The sensitivity of the model was tested against different user preferences. For this purpose three different scenarios were proposed to evaluate the model performances in finding optimum solutions according to requirements imposed by the user and/or planner and/or designer.

Scenario 1

Impose high fill unit cost when an area lacks fill material or the soil properties are not suitable for using as fill material.

In this case the optimum solution must produce higher cut quantities than fill quantities to reduce the total cost. To test the sensitivity of the model against high fill unit costs, a cost ratio of 3:1 (fill unit cost : cut unit cost) was applied and the results were as in Table 6.7 and Figures 6.23 to 6.25. Note that this test was produced using the parameters shown in Table 6.1 with the application of R&P at generation 25.

Table 6.7: Earthwork results for high unit fill cost

Fitness value	10,358,100
Cut quantities	272,975 m^3
Fill Quantities	49,502 m^3
Total Cut – Fill volume	322,477 m^3
Difference in Cut and Fill	272,975 m^3
% of Cut and Fill difference from the total earthwork amount	69.3%

The results show that the percentage difference of cut and fill to the total amount of earthwork is 69.3% for the sake of cut soils. This indicates that the solution algorithms try to find a solution that requires less fill material to reduce the total cost. This is graphically shown in Figures 6.23 to 6.25. This proves the sensitivity of the model against high fill cost values.

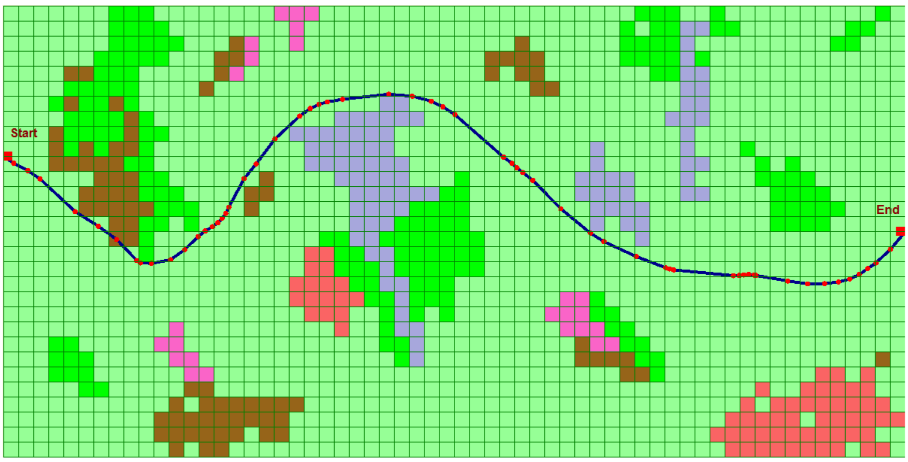


Figure 6.23: The optimum horizontal alignment with high fill unit cost (land use grid map)

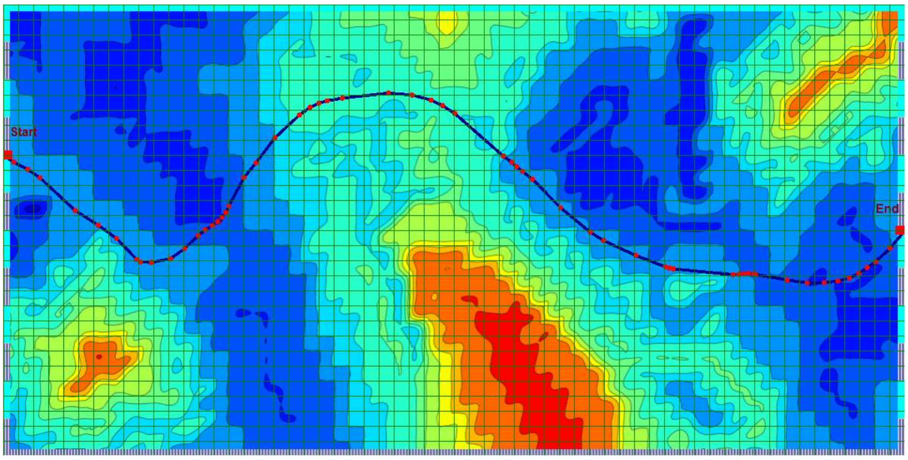


Figure 6.24: The optimum horizontal alignment with high fill unit cost (contour map)

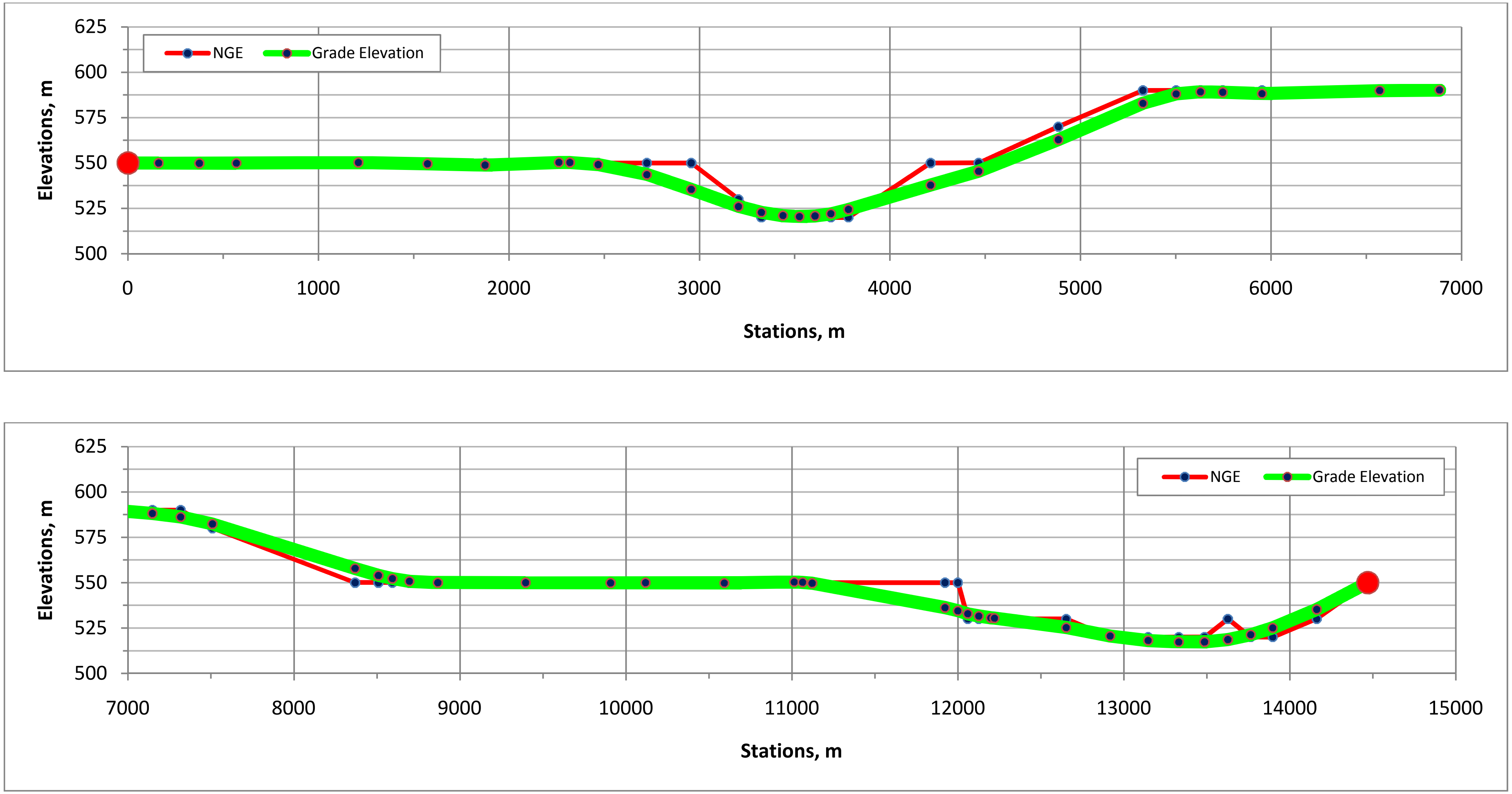


Figure 6.25: The highway profile result with high fill unit cost

Another test was performed using zero cut unit cost. This scenario was supposed to produce a solution with no or minimal fill material. The test result was a solution with 5,159,368 m3 cut volume and only 20 m3 fill volume. The solution alignments are presented in Figures 6.26 and 6.27. From Figure 6.26 it can be seen that the alignment passed through the very high elevation areas (the red coloured area) recognising that these areas incur no cut costs besides a shorter alignment.

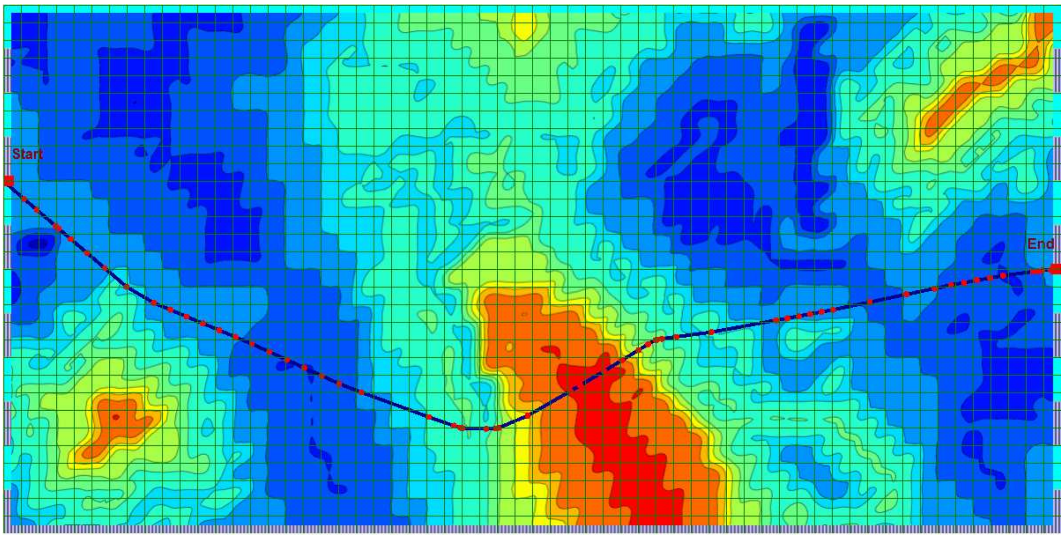


Figure 6.26: An alignment solution with zero cut costs

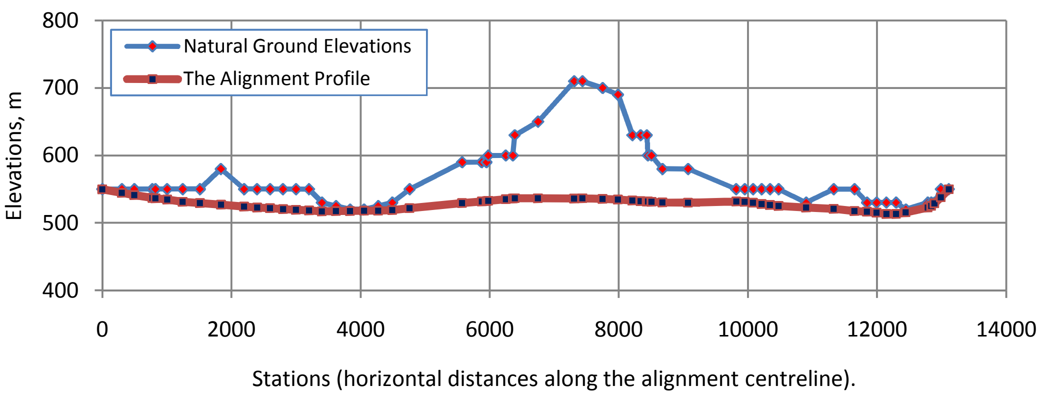


Figure 6.27: The profile of an alignment solution with zero cut costs

Scenario 2

Impose high curvature costs to produce solutions with few or no horizontal and vertical curvature violations. This criterion is important for safe driving conditions.

In Table 6.5, where no special preferences were given for curvature satisfaction, it is shown that (NO VHC) and (NO VVC) of test scenario G1000 R&P@25 were 5 and 6 respectively. This means that the horizontal curvature is violated at 5 and vertical curvature at 6 station points.

The aforementioned test was made using unit horizontal penalty cost (UHPC) and unit vertical penalty cost (UVPC) equal to (1). To test the sensitivity of the model against these parameters, high unit penalty costs were applied and Figures 6.28–6.30 and Table 6.8 show the results:

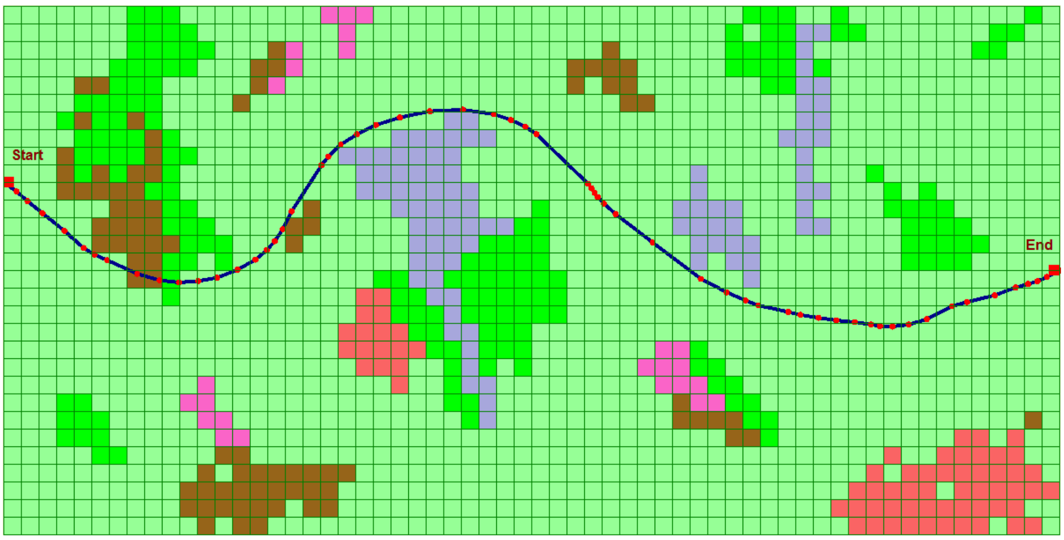


Figure 6.28: A smoother alignment due to higher curvature penalty costs (land use map)

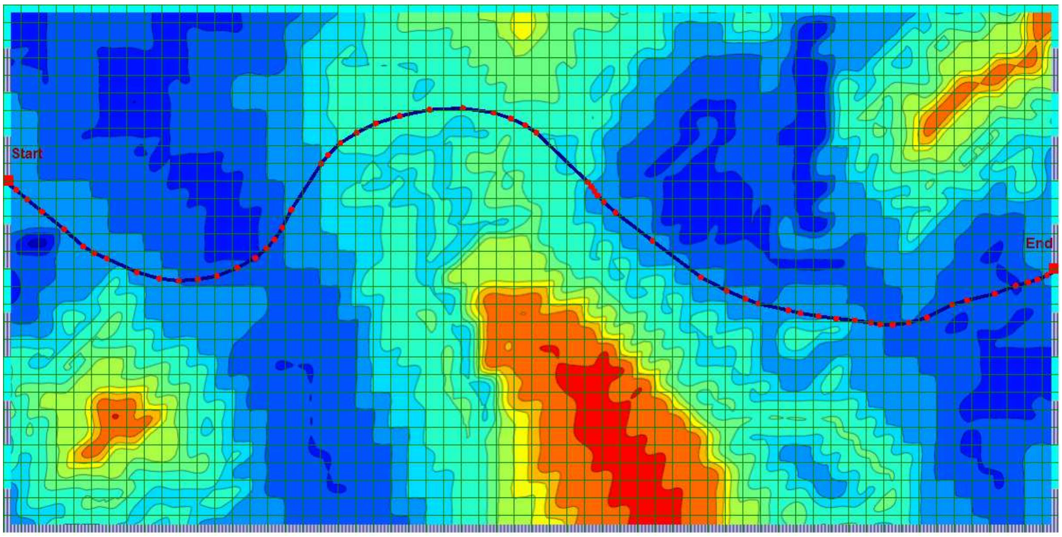


Figure 6.29: A smoother alignment under the effect of higher penalty costs (contour map)

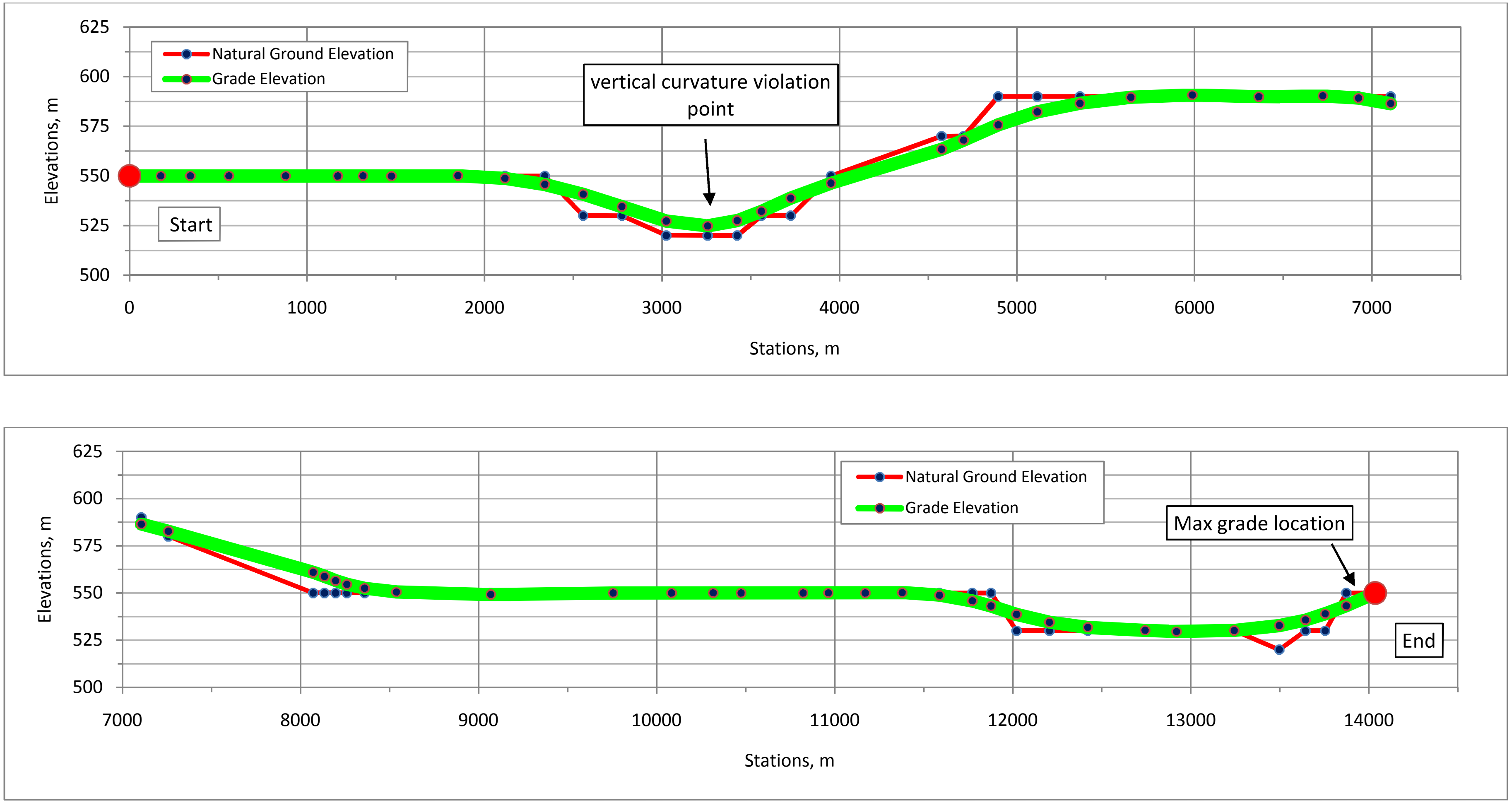


Figure 6.30: A smoother profile under the effect of higher curvature penalty costs

Table 6.8: Curvature results with high curvature penalty costs

Fitness Value	10,039,491 unit cost
Cut volume	130,224 m^3
Fill volume	187,438 m^3
Total Cut – Fill volume	317,662 m^3
Difference in Cut and Fill	57,214 m^3
% of Cut and Fill difference from the total earthwork amount	18%
Exist G_{max}	4.18%
No VHC	0
No VVC	1
THC	360.68 degree
HCI	0.0257 degree/unit length
TVC	35.66 degree
VCI	0.00254 degree/unit length

It can be seen that the solution has HCI equal to 0.0257, which is smoother than the optimum solution found earlier (refer to Figures 6.20-6.22 and Table 6.6) with HCI 0.0283. Moreover, the VCI is also decreased from 0.0027 to 0.00254, thus providing a smoother vertical transition as well. On the other hand, the results show that no horizontal curvature violation is recorded with only one vertical curvature violation, which is located at station point 14 (around station 3300m) (see Figure 6.30). This test proved that the model was sensitive against different curvature weights for smoother horizontal and vertical alignment curvatures.

Scenario 3

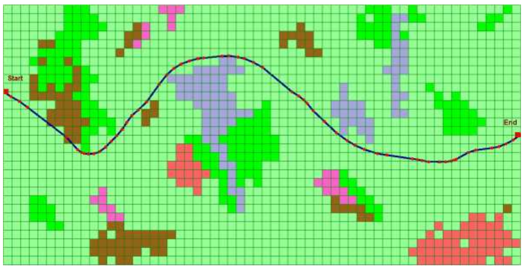
The sensitivity of the model against an additional cost component was tested. For this purpose a user cost function was formulated and added up to the fitness function. The user cost was formulated as follows:

$$C_{User} =$$

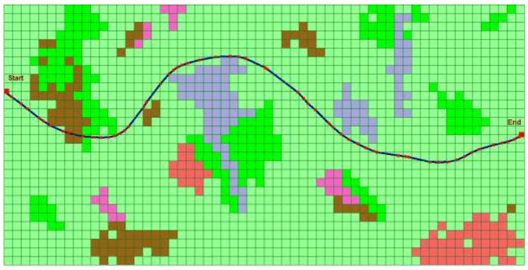
$$\sum_{yr=0}^{D_{yr}} [(TV + TV * ATGR * yr) * (UTC + UTC * APIR * yr) * L - (NP + NP * AProIR * yr) * TV * L] \quad (6.30)$$

This cost (C_{User}) represents the cost incurred by the users to travel along the road, less profit. Here it is calculated for the whole design life (D_{yr}) of the facility which considers the annual traffic growth rate (ATGR), annual price increase rate (APIR), net profit (NP), and annual profit increase rate (AProIR). Then the net traffic cost is calculated by summing up the yearly net user costs by multiplying the traffic volume (TV) by unit traffic cost per vehicle per unit length of the facility (UTC) and by the length of the facility, and then to subtract the net profit that the facility generates.

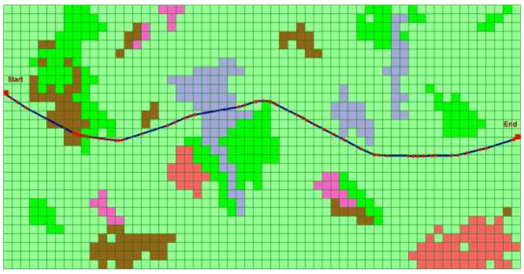
The model sensitivity against user costs was tested using three different design lives and the results compared with the solution which was obtained when no user costs were used. The solutions were obtained using 5, 10, and 20 years design lives. The user cost as per the formulation above tends to reduce the total alignment length to decrease the user costs. Although the different location, construction, and user unit costs and their ratios have different influences on the shape of the final solution, the different alignment configurations that were obtained by the application of different design lives proved the sensitivity of the model to different cost components. Figures 6.31 and 6.32 and Table 6.9 show the results.



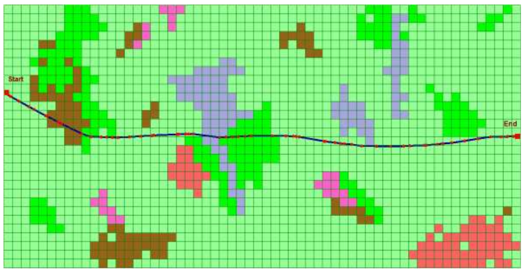
a) No user cost consideration



b) User costs for a design life of 5 years



c) User costs for a design life of 10 years



d) User costs for a design life of 20 years

Figure 6.31: The effect of user costs and different design lives on highway alignment solution

Table 6.9: The effect of different design lives for user cost consideration

User Cost Scenarios	Alignment lengths (m)	Alignment Fitness (unit costs)
No user cost is applied (Case 1)	14,330	8,415,960
User costs for 5 years design life (Case 2)	13,940	35,794,014
User costs for 10 years design life (Case 3)	12,735	82,705,608
User costs for 20 years design life (Case 4)	12,279	245,574,722

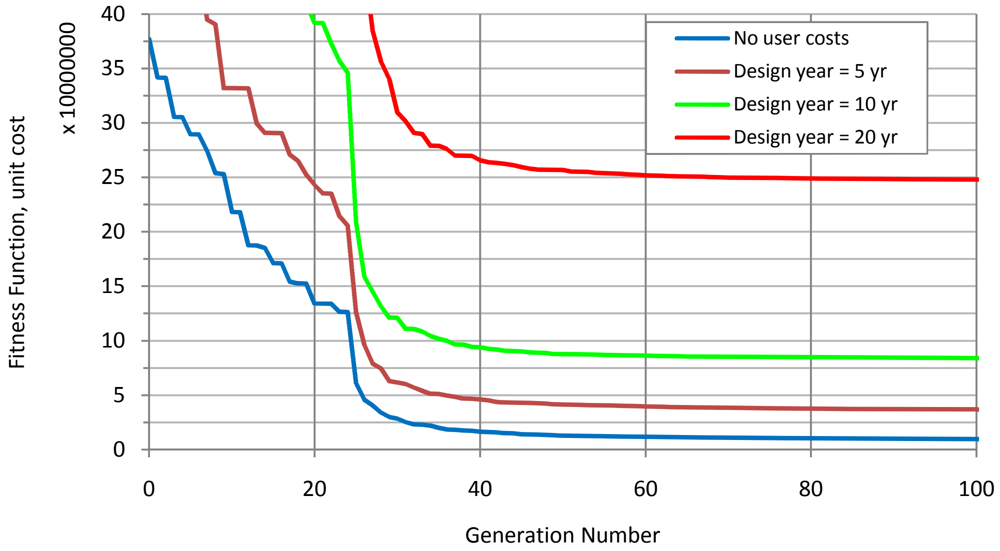


Figure 6.32: Fitness for different design year tests

The results show that high user costs result in shorter alignment. The different results obtained by the solution algorithms increase the credibility of the model for finding good solutions under different circumstances. It should be noted that the proportion of the fitness function's cost components plays an important role in the shape of the final solution and the user must find and apply the unit cost figures that are dependable in real life applications.

6.10.2 VERIFICATION TESTS

To verify the efficiency of the solution algorithms, another case study was prepared and the model was tested upon it. The world of this case study was randomly generated (with dimensions 12,000mX6,000m and grid cell sizes 200mX200m) for both the land use areas and the topography of the area. The test results were as in Figures 6.33–6.35 and Table 6.10.

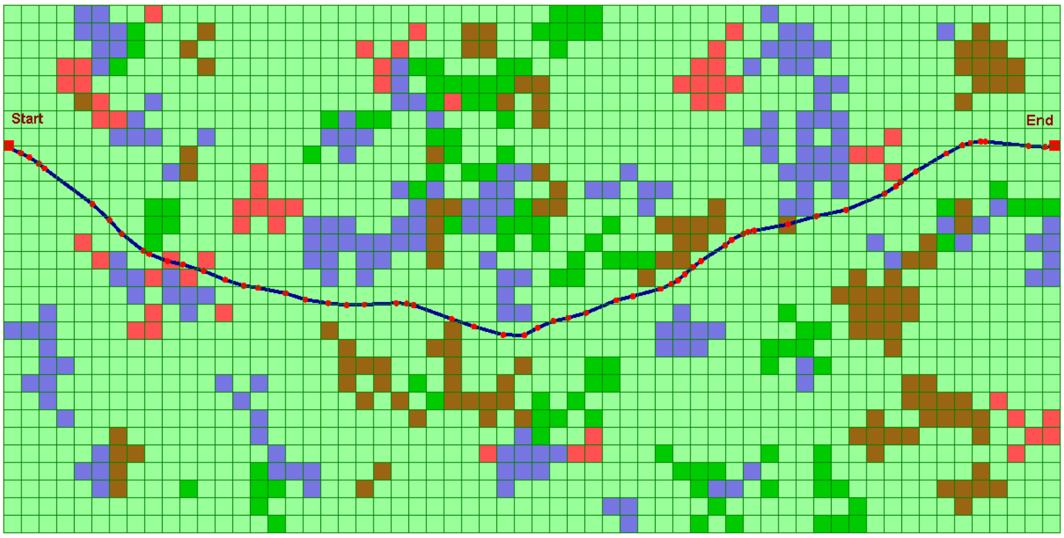


Figure 6.33: An optimum 2D solution result on a random world (land use map)

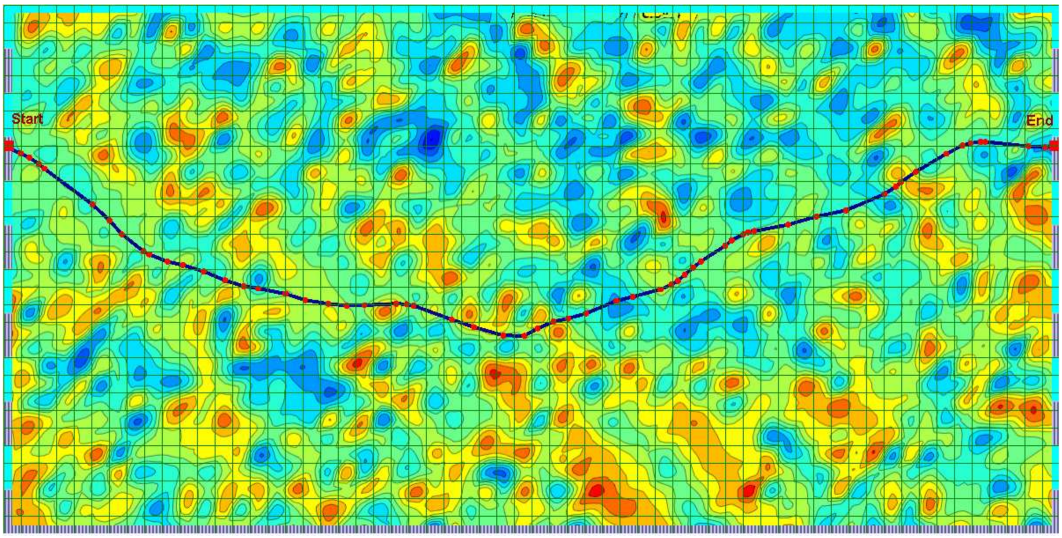


Figure 6.34: An optimum 2D solution result on a random world (contour map)

(The ground elevations of the study area were randomly generated to make a complex land elevation configuration)

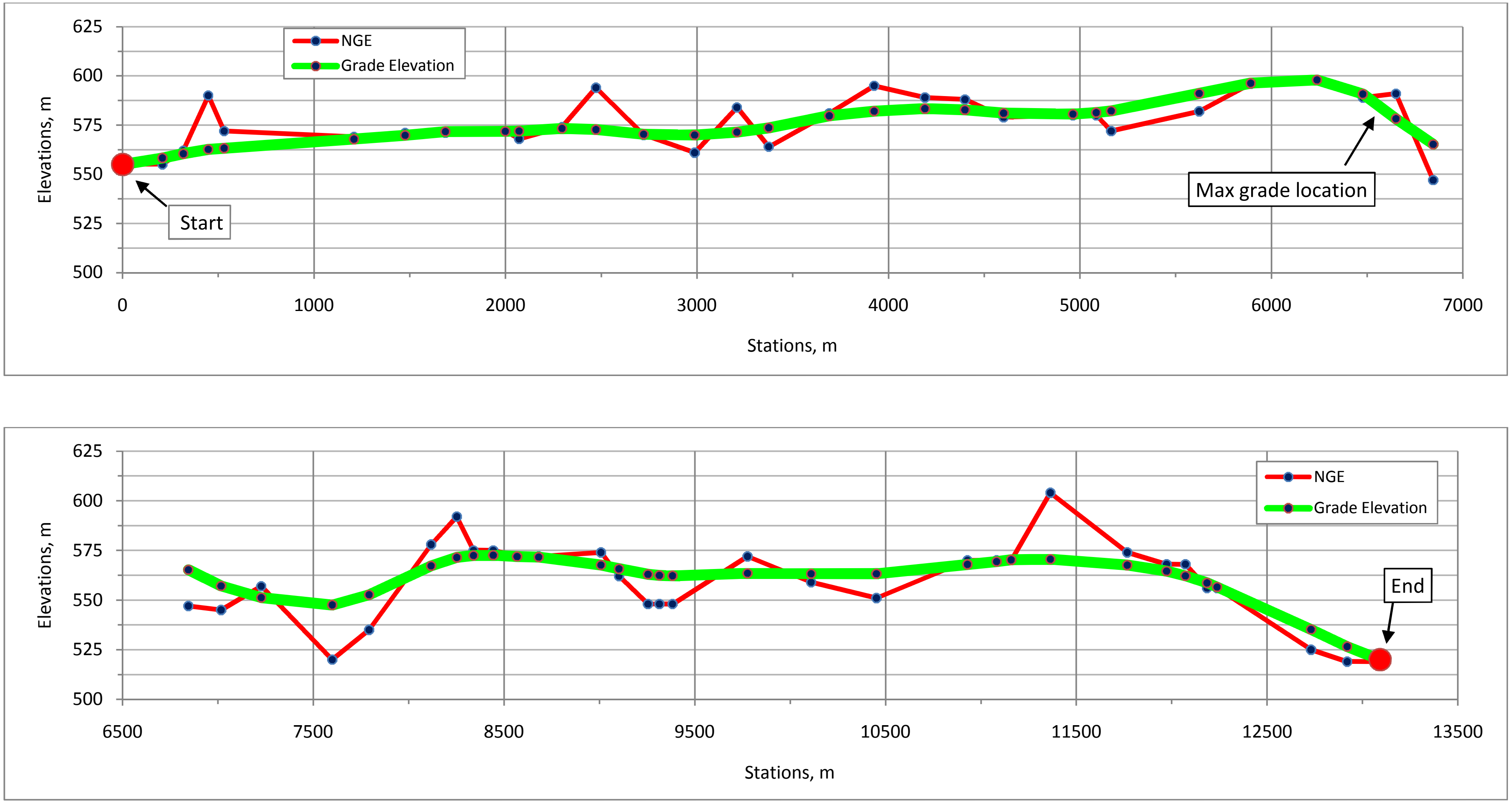


Figure 6.35: The profile

Table 6.10: The results of the alignment characteristics

Fitness Value	14,700,569 unit cost
Cut volume	468,981 m^3
Fill volume	454,464 m^3
Total Cut – Fill volume	923,445 m^3
Difference in Cut and Fill	14,517 m^3
% of Cut and Fill difference from the total earthwork amount	1.57 %
Exist G_{max}	7.26 %
THC	396.965 degree
HCI	0.0303 degree/unit length
TVC	51.9 degree
VCI	0.00396 degree/unit length

It can be seen that the percentage difference of cut and fill is 1.57%, the maximum grade is 7.26% which is less than the maximum allowable grade limit (8%) (see Table 6.5), and the solution possesses smooth horizontal and vertical curvatures. These verify that the model is efficient to produce good solutions with different world problems.

6.11 SUMMARY

The model developed in Chapter 5 for 3D highway alignment is extended in this chapter so that the constraints that are considered important for realistic highway alignment are imposed on the alignment solutions. This was done so that the possibility of handling the constraints by the proposed design approach is tested and also to investigate obtaining a realistic 3D highway alignment configuration that conforms to the standard design policies required for safe and comfortable driving conditions.

The model was formulated to handle horizontal curvature, vertical curvature, and maximum vertical gradient as constraints. The requirement for balancing the quantities of cut and fill earthwork, as a user-defined option, was also embedded into the model formulation. Moreover, the sensitivity of the model against different desirability and project requirement was also tested.

It was assumed that each three successive station points along the alignment form an independent alignment case and their requirement for horizontal curvature, vertical curvature, and maximum gradient was tested. This was important in order to check any point along the alignment against the constraints assigned to the problem. The horizontal curvature was defined using chord definition principles and it was assumed that the distance between the station points are distances representing the chord length necessary to define the required curvature. This was also the case when vertical curvature was configured. It should be mentioned that these constraints were considered using soft penalty and repair techniques.

One of the limitations of the model is that the user should estimate a proper value for the applied penalty unit cost so that it can compromise with the other cost component included in the fitness function; otherwise, the solution would be optimum with curvatures and/or gradients violated at some specific station point locations.

The results showed that the ways in which the defined constraints were handled are valid and realistic alignments could be obtained. Moreover, the model showed a high level of sensitivities against different requirements besides the satisfaction of balancing cut and fill amounts when it is desired by the planner.

This study also suggests that access provision in terms of junctions is an important issue for the land use centres. This may turn the problem from a single highway alignment optimisation to a road network system problem. The location of the junctions, access costs, proximity costs, and the related constraints are all factors that may affect the alignment configuration, location, and total costs. The effects of access provisions and the possibility of optimising junction-highway alignment simultaneously were investigated. The details are discussed in the next chapter.

CHAPTER 7

JUNCTIONS AND HIGHWAY ALIGNMENT OPTIMISATION

7.1 INTRODUCTION

So far, a GA-based model for 3D highway alignment optimisation, through the proposed new approach of station points, has been built. The model was tested against several cost components, constraints, and world cases and proved successful. The different cost, penalty, and constraint components of the model showed critical impacts on the alignment location, configurations, and total fitness value. However, other factors like access provision for the land use centres to the main alignment, which has not been dealt with, may also have an influence on the optimum solution being found. Land use centres within the study area, as areas for traffic generation, need to have access to the main highway alignments. Perhaps accesses are provided in the form of junctions, and therefore optimum junction locations would be inevitable.

This chapter investigates the possibility of using the proposed approach of station points to handle the problem of access provision and to find optimum locations for the junctions that best serve the land use centres (finding best locations for the junctions on the main alignment). Moreover, the study tries to address the effect of the optimum junction locations on the main alignment location and configuration. Thus, the aim is to handle the problem of access provision as junctions with the previously developed model and to find whether or not the junction-highway alignment problem needs simultaneous optimisation. In this study, a preliminary investigation is made on a 2D highway alignment and, if the argument is proved significant, an extensive investigation would be demanded and recommended.

The chapter starts with an introduction to the possible effects of access provision on highway alignments. It also introduces the way that the junctions are represented in the chromosomes with the notion of station points; then it explains the GA operators that are used; and gives some details to the related cost elements. Moreover, the chapter presents some experiments with their results.

7.2 THE PROPOSED APPROACH OF STATION POINTS AND JUNCTION LOCATIONS

The notion of addressing the effect of access provision and junction locations on highway alignment optimisation was inspired by the fact that an important land use centre (e.g. a business or a community centre) with a relatively high traffic volume may pull a highway alignment towards itself to reduce access costs. This could be the case when the independent costs of access and the main alignment incur more costs than if they are considered simultaneously. Figure 7.1 explains that when the alignment is considered with no junction the solution would be alignment A1. When an access link L1 is added to the optimum solution A1, the total cost would be the sum of A1+L1. Meanwhile, when the alignment is simultaneously taken with the requirement for access, the solution becomes A2+L2 and the total costs of A2+L2 would be less than A1+L1. This explanation calls for simultaneous junction-alignment optimisation investigations.

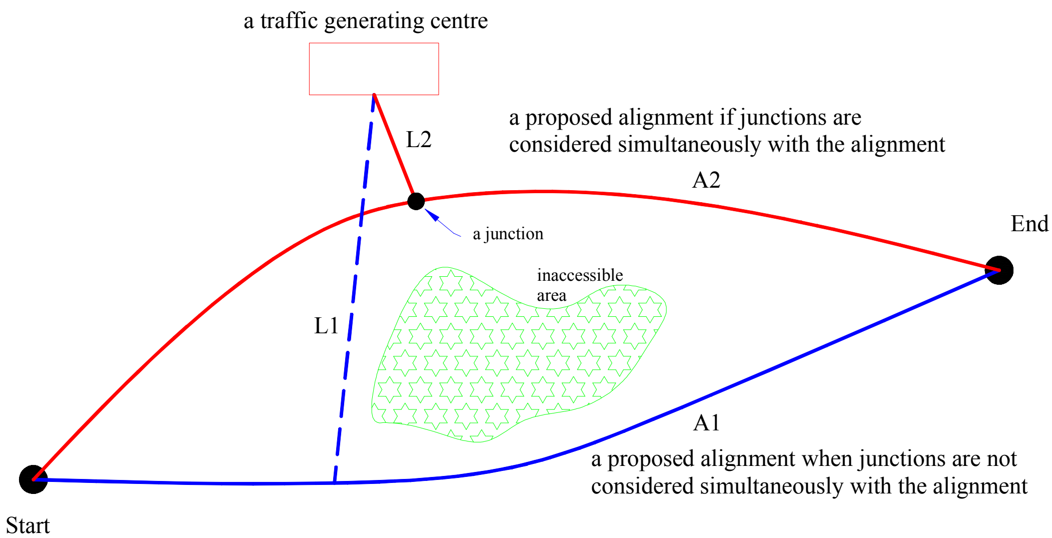


Figure 7.1: An illustration for a possible alignment location with and without junction consideration

The notion of station points for highway alignment configuration, as introduced in this research study, gives flexible ranges to the junction locations. Bearing in mind the station point definition, each station point is represented by their X, Y, and Z coordinates and therefore it is thought that each station point can be used as a probable location for a junction. Thereafter, different alignments with different junction locations are evaluated and a solution with the least fitness is considered optimum.

The sections below describe the required formulation to build a model that integrates the junction problem with the model for highway alignment optimisation.

7.3 GA FORMULATIONS

7.3.1 THE CHROMOSOME AND JUNCTION REPRESENTATION

Figure 7.2 represents the chromosome map for junction representation along a highway alignment.

Index (i)	0	1	2	…	i	…	N-2	N-1
Junction	0	0	1	0	1	0	0	0
Individual (j)	X_0	X_1	X_2		X_i		X_{n-1}	X_n
	Y_0	Y_1	Y_2		Y_i		Y_{n-1}	Y_n

Figure 7.2: Chromosome representation for junctions and 2D highway alignment

It can be seen that the representation is the same as for 2D highway alignment with the addition of a row for junction representation. The junctions were represented in a binary format (0 and 1) where 0 means no junction at the specified location, while 1 means that a junction exists. In this research, the determination of the number of the junctions (nj) was made a user-defined parameter. The number of the traffic generation centre (land use centres) and their locations may have influence on the number of the junctions being provided.

The initial population is generated using the algorithm shown in Figure 7.3.

Let nj be the total number of junctions that are being provided, N the total number of station points, and $J_{j_{no}}$ the random junction location. For each individual generate random locations for the junctions:

For j_{no} = 1 to nj

$J_{j_{no}} = rnd(1, N-1)$

Next j_{no}

Figure 7.3: An algorithm for random junction location generation

This algorithm finds the order or index of the station point, between 1 and N-1, as a possible location for the junction, where this location is defined by the X and Y coordinate of the station point.

Now a gene (i) of a chromosome constitutes of a possible junction location, X coordinate, and Y coordinate of the station point.

7.3.2 CROSSOVER OPERATOR

Two approaches were suggested to apply the crossover operator on the chromosomes.

Approach 1:

Use the same RMPCrO crossover as for the 2D highway alignment.

This method may lead to produce offspring with an invalid number of junctions (different than *nj* which is initially set or defined by the user or planner). Greater or fewer numbers of junctions may be generated on the resulted offspring and when these happen the chromosome needs to be repaired either to bring back the lost junctions or to remove the extra ones. A repair algorithm was developed and applied on the chromosomes. The algorithm checks the number of the junctions on the offspring chromosomes and, if it is different than *nj*, it will be repaired. The method of chromosome repairing is done on random bases and thus the effect of crossover is either reduced or dismissed.

Approach 2:

Apply mutation only on the junctions and prevent them undergoing crossover.

This was suggested under the assumption that the new location exploration for *nj* number of junctions might not be so critical (sensitive) to the GA operators and therefore a suggestion was made to handle them by mutation only. Moreover, the result of the offspring chromosome creation with approach 1 above results in a similar effect as if approach 2 alone is used.

7.3.3 MUTATION OPERATOR

A standard uniform mutation, as for binary representation, was applied on the genes that were specified to handle positions for the junctions. The mutation method randomly selected a junction position and another non-junction position and flipped their components over as follows:

Select a junction position randomly (1) and turn its value to 0.

Select a non-junction position randomly (0) and turn its value to 1.

7.4 JUNCTION RELATED COSTS

This research dealt with junction optimisation as an attempt to investigate the effect of access provisions as junctions on the highway alignment location and configuration. It was assumed that the distances between the land use centres and the junctions on the main alignment are direct and straight and the junctions' related cost components were formulated based on this assumption. Future works can deal with the real characteristics of the study area and make an inclusion for the requirement of a winding link between the land use centres and the junctions.

The following cost components are formulated based on the assumption of a straight link.

7.4.1 ACCESS COST

The access cost of this study was associated with two cost categories:

a) traffic or user costs from the land use centre(s) to the junction(s);

b) construction cost of the link from the land use centre(s) to the junction(s).

For each land use centre (A_i) the shortest distance, when there was more than one junction, between the land use centre and the junctions $(L_{i\_min}\ (A_i\, J_k))$ was determined (Figure 7.4) (where i represents the order of the land use centres and k represents the order of the junctions on the alignment), and then the user and construction costs were found, as in equations 7.1 and 7.2:

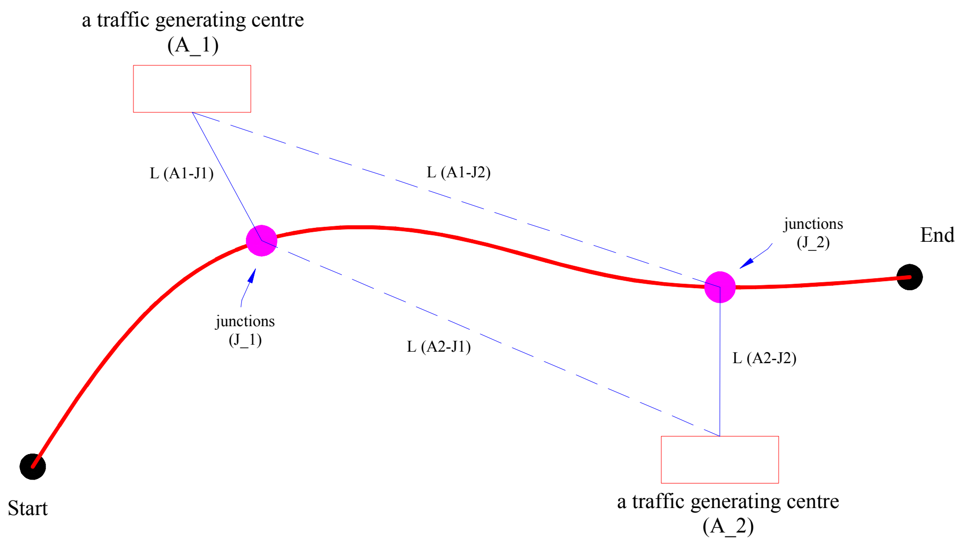


Figure 7.4: The relation of junction location and the alignment configuration and location

$$UC_{Link} = \sum_{i=1}^{i=p} L_{i\_min} * TVA_i * UTC_{Link} * D_{yr} \tag{7.1}$$

$$LC_{Link} = \sum_{i=1}^{i=p} L_{i\_min} * ULC_{Link} \tag{7.2}$$

Where UC_{Link} is the total user cost of the links between the land use centres and the junctions, LC_{Link} is the total length (construction) cost of the links, L_{i_min} is the minimum distance of the i^{th} land use centre to the nearest junction, TVA_i is the total traffic volume that uses the link between the land use centre A_i and the junction on the main highway, UTC_{Link} is the unit traffic cost per vehicle per unit length of the link, ULC_{Link} is the unit length (construction) cost of the links, D_{yr} is the design life, and p is the number of land use centres.

It should be noted that the minimum distance (L_{i_min}) was considered under the assumption that the user tends to select the shortest link to get access to the main alignment. Therefore, when the distance of a junction was located as the shortest, the whole traffic volume of the link was directed to that junction and then considering that the other junctions are not available for this land use.

7.4.2 PROXIMITY COSTS

Proximity is the disadvantage value of a highway alignment for being close to a land use centre. Traffic noise and air pollution are among the factors that an alignment may impose on the areas due to their close distances to the alignment. Considering sustainable development elements, an alignment must not be located as close as the distance that is specified by the planner. In this study this cost was formulated in the form of a soft penalty function in which the alignment was penalised when its location at any point along the alignment had a distance smaller than that specified by the planner, so:

$$if\ L_i < UDL_{min}$$
$$for\ i = 0,1,2, \ldots, N-1$$
$$where\ N\ is\ the\ total\ number\ of\ the\ station\ points, then:$$

$$Prox_{cost} = (UDL_{min} - L_i) * UDProx_{cost} \quad (7.3)$$

Where $Prox_{cost}$ is the total proximity cost that an alignment generates due to its close location to a land use centre, $UDProx_{coost}$ is the user or planner defined proximity unit cost per unit distance of the land use centre and the alignment, L_i is the actual calculated distance between the land use centre and the alignment at any point, and UDL_{min} is the minimum user defined distance between the alignment and the land use centre.

7.4.3 THE FITNESS FUNCTION

The above access and proximity costs were added up to the fitness function of the 2D highway alignment model, so that:

$$Total\ Fitness\ Function\ is\ to\ minimise = a_1.C_{Length} + a_2.C_{Location} + a_3 * PC_{H\_Curvature} + a_4 * UC_{Link} + a_5 * LC_{Link} + a_6 * Prox_{cost} \quad (7.4)$$

7.5 EXPERIMENTAL TESTS AND THE RESULTS

The experiments were set to investigate the following as the main objectives.

- The possibility of integrating the access provision problems with the proposed approach of station points.
- To optimise junction locations on highway alignment.

- To address the effect of optimum junction locations on highway alignment location and configuration.
- To investigate the possibility of junction-alignment optimisation simultaneously.

The parameters that have been found for optimum smooth 3D highway alignments (Table 6.1 in the previous chapter) were used to perform the tests. The only difference with the test parameters was that the parameters that were related to the third dimension (Z) were switched off to clearly show how junctions affect highway alignments. Several test scenarios were prepared and were divided on two groups based on proximity cost considerations. These were designed so that the effect of each component can clearly be investigated. The following are the different test scenarios with the results.

Group A: With no proximity cost considerations

Scenario A-1:

Where no land use centres was considered and no junction was provided (Table 7.1).

The result was as in Figure 7.5. The alignment solution was smooth and satisfied the horizontal curvature requirements, was as straight as possible, was the shortest possible location, and skirt around the high cost areas.

Table 7.1: Test parameters for no junction consideration (scenario 1)

World	P_{size}	CrO	SS	MM – main alignment	G
W_9	5000	RMPCrO-8P	RndSS	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	1000
No of Land Use Centres	No of Junctions	Chord Length	R_{min}	Start Repair and Penalise Horizontal Curvature at generation (g):	
Nil	Nil	30	200	50	

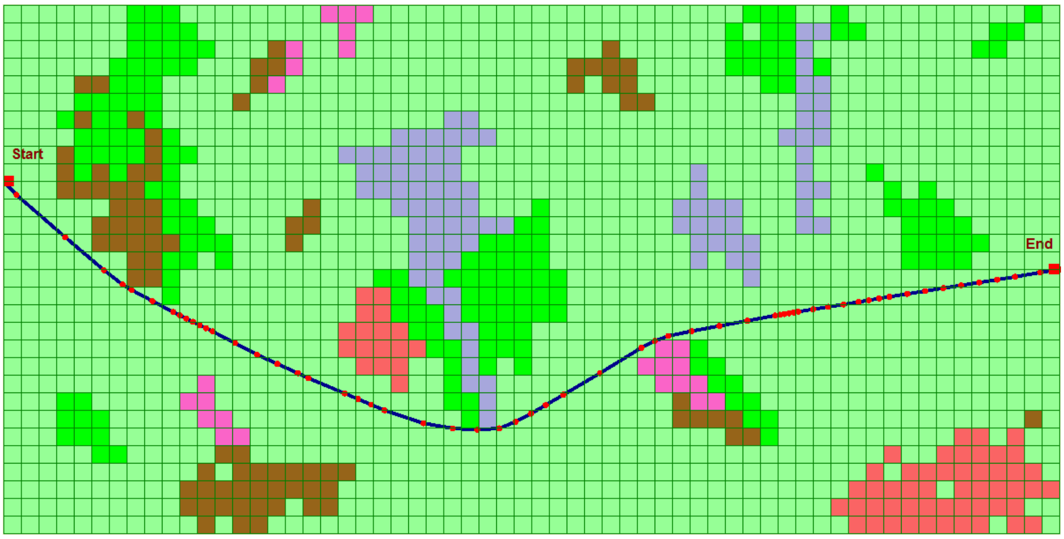


Figure 7.5: An optimum 2D solution with no land use centres and no junction consideration

Scenario A-2:

Where only one land use centre was considered with 800 vehicles/day and one junction was provided (Table 7.2).

The result was as in Figure 7.6. The alignment solution was smooth, was as straight as possible, was the shortest possible location, skirt around the high cost areas, passed close to the land use centre (the yellow rectangle), and provided a junction on the main alignment which was as close as possible to the land use area (the blue circle on the main alignment) to reduce access costs.

Table 7.2: Test parameters for one junction consideration (scenario 2)

World	P_{size}	CrO	SS	MM main alignment	MM junction	G
W_9	5000	RMPCrO 8	RndSS	(2/3)GPM (10% PMR) + (1/3)SM (3% PMR) (UEA: 1/1)	SM	1000
No of Land Use Centres	No of Junctions	Chord Length	R	Start Repair and Penalise H Curvature at generation (g):	Link Traffic vpd	
1	1	30	200	50	800	

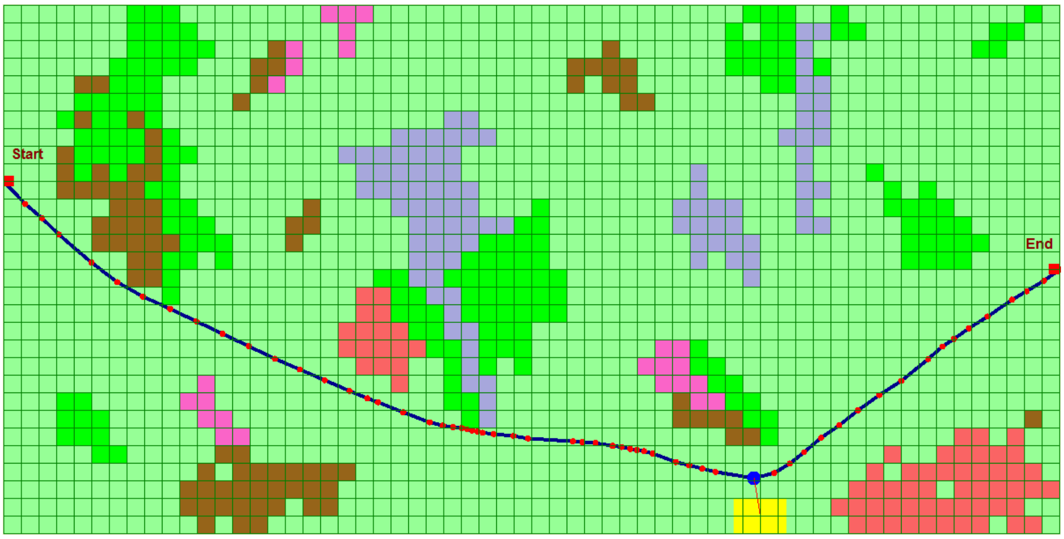


Figure 7.6: An optimum 2D solution with one land use centre and provision of a junction

Scenario A-3:

Same as for scenario 2 but two other different locations were considered for the land use centre with different traffic volumes.

It can be seen that the locations of the main alignments were significantly affected by the location of the traffic generation centre (land use centre) (see Figures 7.7 and 7.8). 1,000 and 2,000 vehicles/day were considered as traffic volumes for the solutions presented in Figures 7.7 and 7.8 respectively. The traffic volumes were chosen by trial and error to make a cost balance between the main alignment and the link costs so that the result is the solution that passes close to the land use centre. This also shows the sensitivity of the model against different traffic volumes.

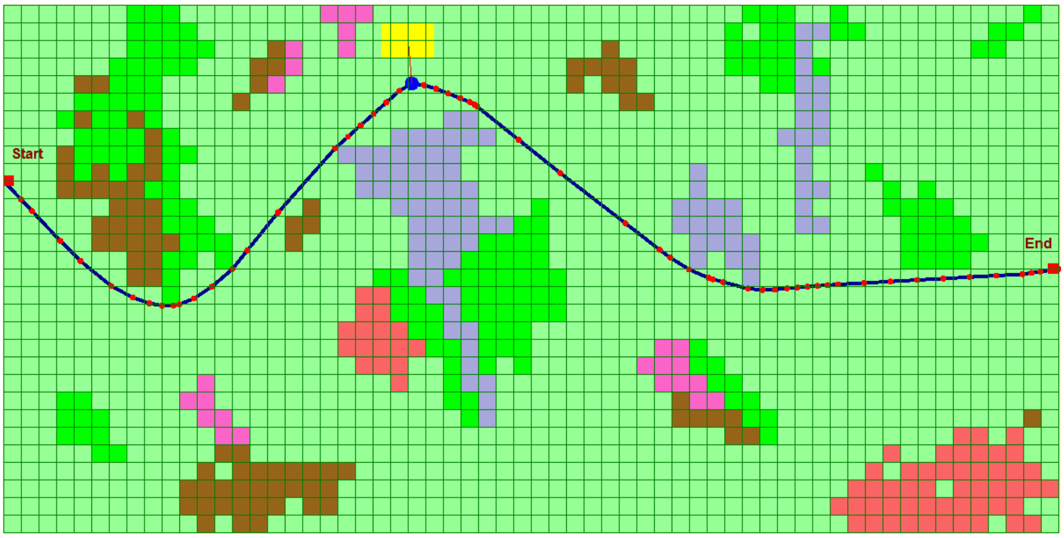


Figure 7.7: An optimum 2D solution with one land use centre at a different location

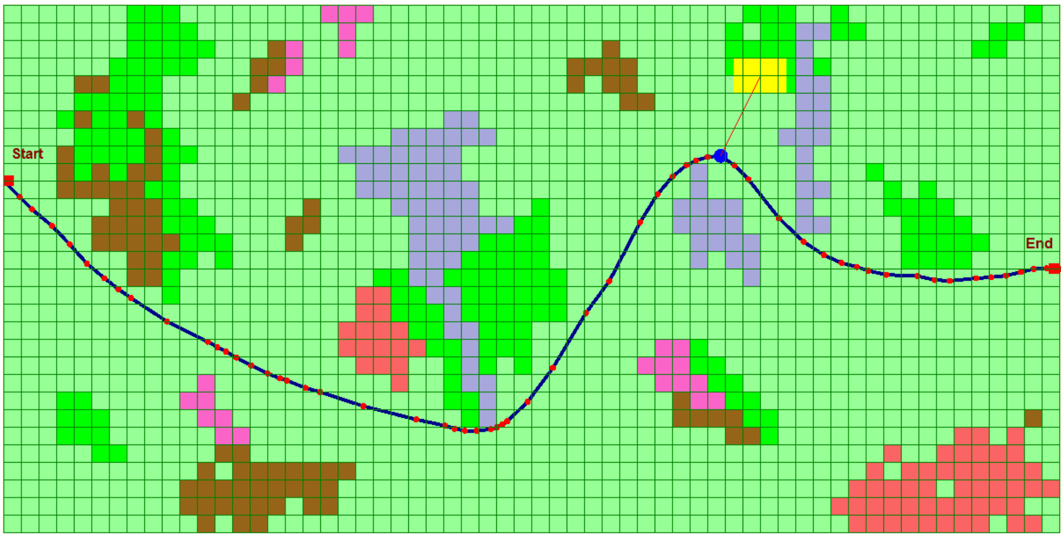


Figure 7.8: An optimum 2D solution with one land use centre at a different location

Scenario A-4:

When two land use centres and the provision of two junctions were considered.

The result was as in Figure 7.9. The alignment solution was smooth, was the shortest possible location, skirt around the high cost areas, passed close to the two land use centres (the two yellow rectangles), and situated the two junctions as close as possible to the land use areas (the blue circles on the main alignment) to reduce access costs. The traffic volumes for these two areas were 1,500 vehicles/day for the one on the left and 2,000 vehicles/day for the one on the right.

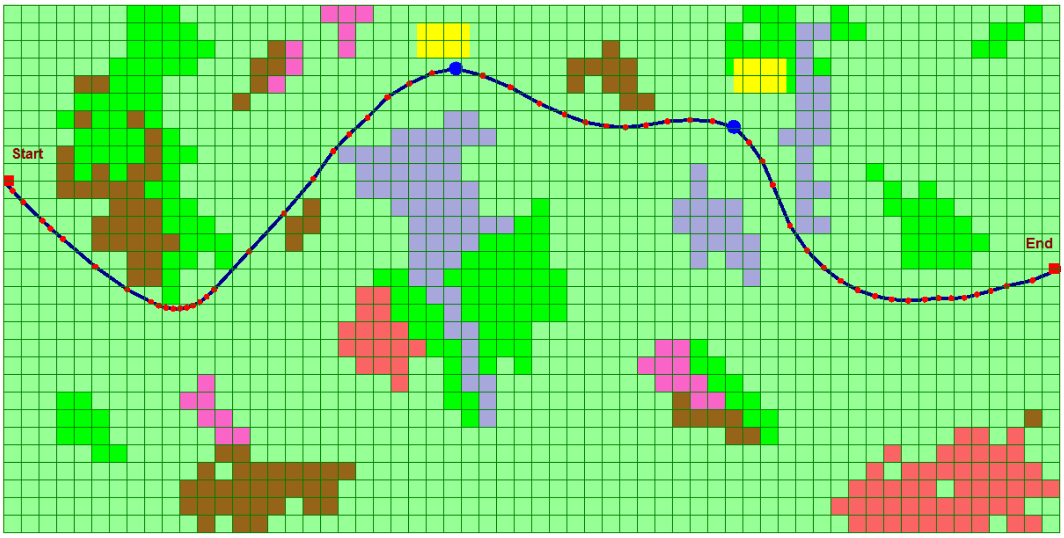


Figure 7.9: An optimum 2D alignment with the provision of two junctions

Scenario A-5:

When three land use centres and provision of three junctions were considered.

The result was as in Figure 7.10. The traffic volumes of the areas were 2,500 vehicles/day for the one on the left, 2,000 vehicles/day for the one in the middle, and 1,500 vehicles/day for the one on the right. The alignment solution was smooth, was the shortest possible location, skirt around the high cost areas, and passed close by the three land use centres where the junctions are provided (the blue circles on the main alignment). It can be seen that the alignment at junction no 1 (the first junction on the left, Figure 7.10) passes through the pink coloured areas and skirts the brown coloured cells due to the fact that the brown cells possess higher costs than the pink ones (refer to Figure 5.28 in Chapter 5 for the coloured cell costs).

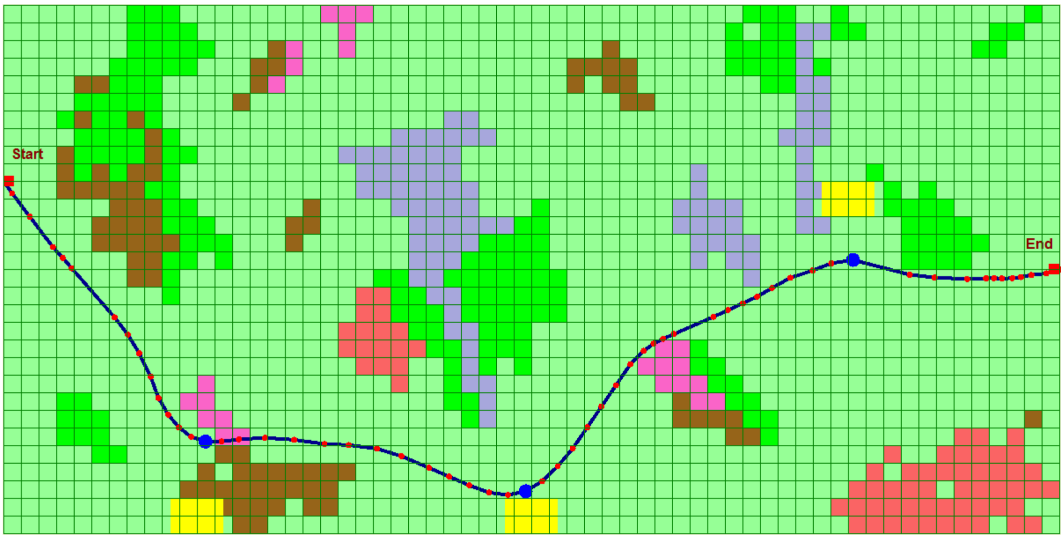


Figure 7.10: An optimum 2D alignment solution with the consideration of three land use centres

The convergences of the above scenarios are shown in Figure 7.11. It shows that the more the land use centres and junction provisions, the slower is the convergence. The model requires further efforts with the three junctions to perform the search compared with no junction consideration or with the scenarios for the provision of one junction.

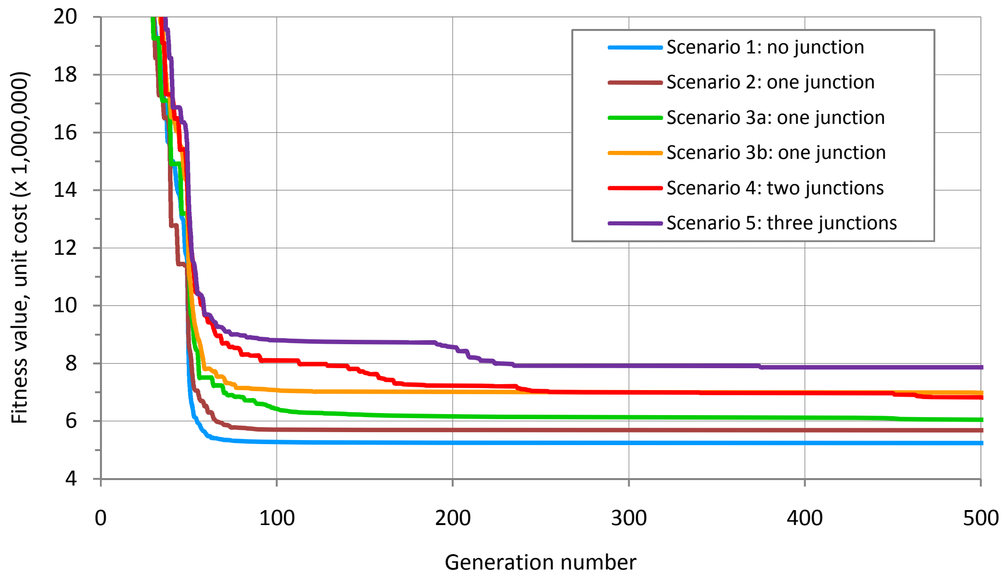


Figure 7.11: Convergence graph of different test scenarios for junction consideration

The results show that access consideration for land use centres has a great impact on highway alignment optimisation. The location and configuration of the main alignment are greatly affected by the land use centre locations when access provision is considered through junctions from these areas to the main alignment.

The above test scenarios also proved the successfulness of the proposed station point approach to handle junctions for access provision. Furthermore, it was proved that an optimum highway alignment cannot be obtained through simultaneous horizontal and vertical alignment alone but also needs consideration of access provision. This notion confirms that simultaneous highway alignment and junctions are inevitable.

Group B: With proximity cost considerations

Scenario B-1:

Two test scenarios were designed with which it was desired to provide a junction for a land use centre with 10,000 vehicles/day.

The first test was carried out with no proximity consideration and the alignment result passed through the land use centre to minimise the relatively huge traffic access cost of 10,000 vehicles/day. The result is shown in Figure 7.12. It can be seen that this solution differs from the alignment obtained with no junction consideration, as shown in Figure 7.5 above.

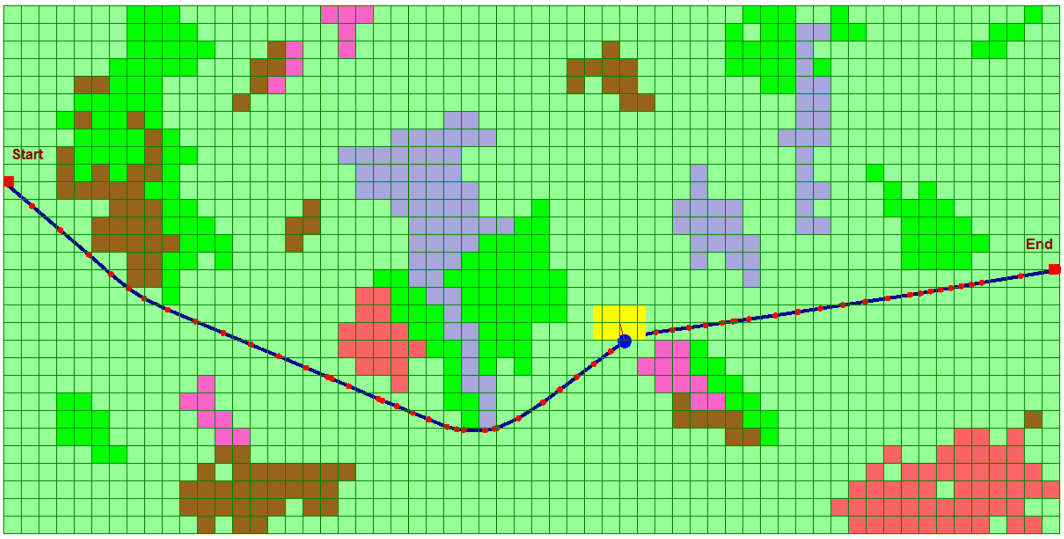


Figure 7.12: An alignment solution with no proximity consideration

The second test was carried out for the same amount of traffic but with the inclusion of proximity consideration. A soft penalty cost of 2,000,000 unit cost per one metre length of violation was applied when the solutions violated the minimum distance, which was defined by the user, between the land use centre and the alignment. In this test scenario this minimum distance was set at 1,000m. The result is presented in Figure 7.13. It can be seen that the alignment passed by more than 1,000m away from the land use centre (more than five cells where each has a dimension of 200x200m) and the junction location was put as close as possible to the land use centre. Figure 7.14 shows the convergence of the two tests.

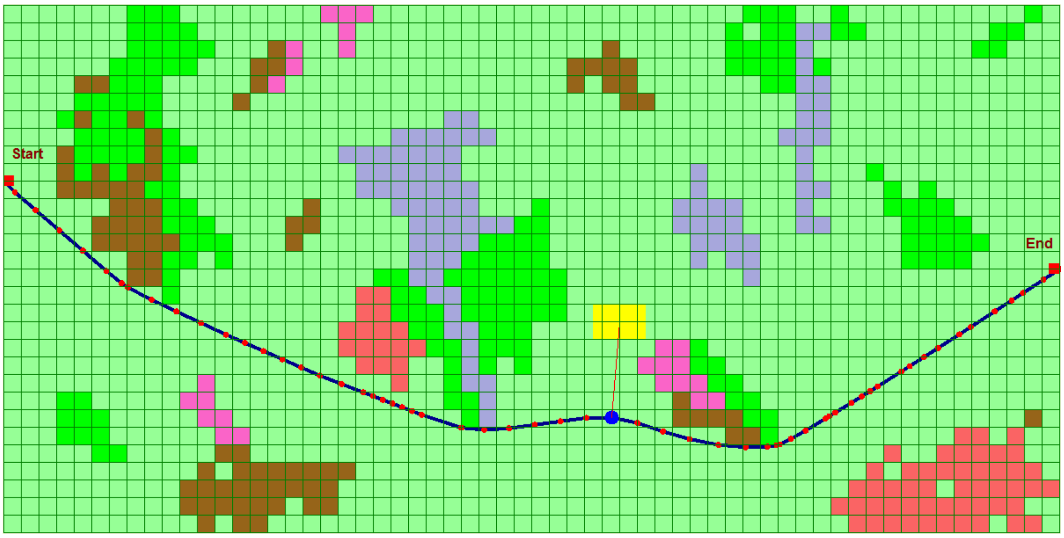


Figure 7.13: An alignment solution with proximity consideration

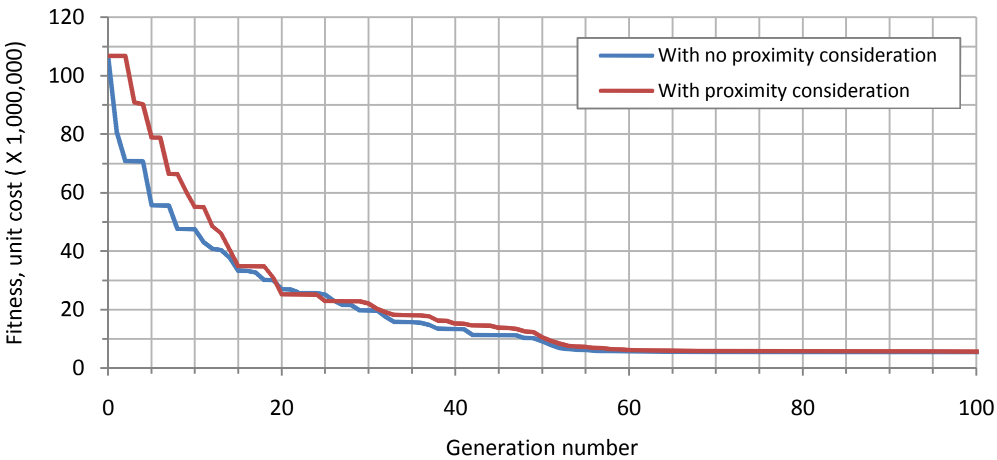


Figure 7.14: Convergence graph of the proximity tests

These tests proved that different access costs including proximity have substantial influences on highway alignment optimisation. Furthermore, these tests proved the sensitivity of the problem against different cost categories.

7.6 SUMMARY

This chapter was dedicated to addressing two main issues. First, preliminary investigation of the possibility of considering access provision as a junction with the proposed approach of station points and finding optimum junction locations; and second, to find the effect of junctions on the main alignment location and configuration. Furthermore, a proximity function as a possible constraint (to protect the environment of the area) defining the minimum distance required between the alignment and the land uses was also formulated. The proximity constraint was considered in terms of a soft penalty function embedded into the main fitness function. The model was formulated in such a way that different judgemental factors reflecting the different aspects of the different land uses and people of interest, including the decision makers, are included. The approach allows interpreting the public/planner/decision maker's opinion into a weighting factor representing their desire. This was to either put the alignment close to a land use or make it far away from it according to the requirement and the purpose of the land use.

The model assumed that the distances between the land uses and the junctions are straight and direct and took the shortest link between a land use and the junctions into consideration for projecting the total traffic and calculating the access and traffic costs on the link.

The model developed in this chapter is unable to consider the main alignment in a 3D space. It deals with the land and topographic feature as being uniform with no substantial change in land and topographic characteristics.

The results show that the notion of the station point approach is successful in handling junctions as locations for access provision from the different land use centres to the main highway alignment. The model was able to find optimum junction locations and showed that junction provision and their locations greatly influence the main alignment location and configuration. Different alignment locations and configurations were obtained with different land use locations. These results strongly recommend simultaneous junction-highway alignment optimisation. It was also shown that a minimum distance can also be defined for better environment consideration for the land uses affected by the main alignment location.

As the idea of junction-highway alignment optimisation has proved successful with the proposed approach of station points, further investigations and formulations are therefore recommended for future studies. This study has opened a wide door for further researches to address the alignment optimisation problem with junctions in further detail. Recommendations are suggested in the next chapter for future works.

CHAPTER 8

CONCLUSIONS AND FURTHER WORK

8.1 INTRODUCTION

This chapter includes a restatement of the objectives and sub-objectives described in Chapter 1. Specific conclusions are drawn from the work and related to the objectives and sub-objectives of the study. The difficulties with the work done are also revealed. Furthermore, the general conclusions that they are not specific to any stated objective and sub-objectives are also drawn. The reliability of the model is also presented through discussing the implications of the developed model in real world applications. The chapter ends with a description of possible areas for future work.

8.2 AIMS AND OBJECTIVES

As stated in Section 1.4, the overall aim of this study was to investigate the possibility of improving the highway alignment development process by the use of modern technologies and, as a sub-aim, to deal with highway alignment as part of a road network system through considering junctions for access provision.

In order to achieve the above-stated aims, the study set the following main objectives:

1. Investigate the possibility of using station points along the centreline of the proposed highway alignment as a new design approach for simultaneous horizontal and vertical alignment development.
2. Ensure that any alignment produced by any model proposed as part of the research satisfies horizontal and vertical curvature directly through the station points' configuration based on predefined radius and K values, respectively. These curvature constraints are important to generate a realistic solution

according to the required design policies for safe and comfortable driving conditions. Limitations for vertical gradients are also included.

3. Investigate the possibility of incorporating access points in the form of junctions with the proposed design approach of station points, and then investigate the effect of access provision and their optimum locations on highway alignments so that the idea of simultaneous junction-highway alignment optimisation is discussed and, if successful, is introduced. This is expected to change the definition of highway alignment optimisation from a single highway alignment problem to a joint junction-highway alignment optimisation problem. The idea of simultaneous junction-highway alignment optimisation would promote a sustainable development of the area under consideration.
4. Investigate the possibility of using GA, as a search tool, for solving the problem of highway alignment development with the use of the proposed design approach.
5. Test the viability of the proposed approach for highway alignment development.

To achieve the above objectives, understanding and identifying the fundamentals of the problem components became essential for introducing the new design approach and to developing a highway alignment. Therefore, the following sub-objectives were also defined:

1. Find out how the problem was dealt with in other researchers' work and identify the main components of a good model.
2. Identify the factors that affect highway alignments.
3. Understand the characteristics of highway alignment geometric design elements and their standard requirements for a realistic alignment and safe driving condition. This is important in order that all relevant characteristics can be incorporated into any approach using station points.
4. Identify the cost components that are thought critical to the problem.
5. Understand the principles of GIS and the ways that it deals with spatial and geographic data. This is thought important to find a proper way to format the

study area in the suggested model. This includes developing a testing regime (worlds and the features that could be included) which reflects the practical situation that might arise in highway alignment planning.

6. Understand the elements of sustainability and consider them for both the main alignment development and when access provisions are considered for the land use centres.
7. Investigate the possibility of formulating and solving the optimisation of a model as a genetic algorithm. In order to do this it will be necessary to investigate:
   - the possible representations of the problem in terms of the physical world and the chromosome of the problem
   - the population size for any particular problem
   - the procedures for selection
   - the procedures for crossover
   - the procedures for mutation
   - the rates for applying mutations.
8. Determine whether the GA can produce optimum and realistic alignment.

It is worth mentioning that the idea of this research study was inspired by technology development; and, also, it was based on the fact that any generated alignment by whatever method will finally consist of a series of station points and it will be implemented (set out) on the ground depending on these station points. Therefore, it was expected that the proposed approach would provide immediate field data for setting out the design on the ground in the same way as providing facilities at the planning and design stages. The design data of the optimum solution, represented as the X, Y, and Z coordinates of the station points, can be uploaded to a total station and the station points' positions are downloaded on the grounds of which the exact alignment design is set out without the need to do additional calculations or techniques as the traditional methods do.

It was stated that if it is proved that station points could be used to determine the alignment then it should be possible to combine the stages of planning, design, and generating the data required for implementation (the station points) in one process, and consequently the development process would be simplified, and the total time and the total cost are significantly reduced.

Each of these main and sub-objectives is now specifically considered in turn, in addition to some other general conclusions. The conclusions are drawn based on the experiments and analysis provided in the chapters of this study.

8.3 THE RESEARCH FINDINGS

8.3.1 THE SPECIFIC CONCLUSIONS TO THE RESEARCH MAIN OBJECTIVES

Objective 1

The method of configuring highway alignments through station points were proved viable through obtaining optimum 2D and 3D highway alignment results from experiments on worlds with absolute known solutions. The results were found to be as close as +99% to the absolute known optimum solution. This was detailed in Chapter 4 Section 4.5.1 for 2D highway alignment and in Chapter 5 Section 5.5.1 for 3D highway alignment. Moreover, the results showed that the proposed approach is capable of handling both horizontal and vertical alignments simultaneously (for details, refer to Chapters 4 and 5). It was shown that different optimum alignment solutions were obtained when they were considered in separate stages. This was shown in Chapter 5 Section 5.5.2.

Objective 2

The experiments described in Chapter 6 proved that the model was successful to produce alignment solutions that obey horizontal curvature, vertical curvature, and maximum gradient requirements. The introduction of the station point approach, as with the traditional (existing) design method, necessitates the satisfaction of the standard design requirements for safe and comfortable driving conditions at any point along the alignment. It was found that the locations of the station points are the key factor in producing a smooth alignment.

From the results, it was proved that the alignment segments between any three successive station points need to be considered as independent alignment cases and need to be checked for horizontal curvature, vertical curvature, and gradient. The above-mentioned criteria were considered as constraints and handled by applying soft penalty and repair techniques. Moreover, the experiments showed that the standard design requirements for the whole alignment could be achieved when the standard requirements are checked and satisfied on the station points one by one.

Objective 3

Although the problem of junctions was not investigated thoroughly, the results gave indications that a highway alignment location cannot be optimum without the consideration of providing access to the areas (the different land uses) that produce traffic and located within the boundary of the study region. The following are considered the main findings for providing access.

- The results proved that the station point approach can handle locations for junctions and the model was able to optimise the junctions' locations along the alignment. The junction locations on the main alignment were greatly affected by the land use locations and the amount of traffic that these areas are producing.
- The results showed that access provision and optimum junction locations have critical impact on the final optimum alignment solution. The location and configuration of the main alignment were greatly affected by the consideration of junctions compared with the results with no junctions. This

concludes that a highway alignment cannot be optimum without considering the junctions and the alignment simultaneously. This could also promote a sustainable development of the area.

The issue of junction problem was covered in detail in Chapter 7.

Objective 4

The GA, as a search tool, proved efficient in exploring the search spaces and locating a good alignment solution. It was found that the efficiency of the search using GA is a function of the GAs' components and operators. The findings about the parameters and the design of the operators are detailed with the conclusions specific to the sub-objectives (Section 8.3.2).

Objective 5

The model, as shown in Chapters 4 and 5, proved viable for highway alignment development. Different alignment configurations and locations were produced when different criteria were applied. The model was tested on several worlds with different land and topographic features. The solutions were realistic, well-located, and able to minimise the associated cost function.

The sensitivity of the model was also tested and it was found that different input parameters make the model respond accordingly. It was proved that different cost items and different weighting factors produce different alignment solutions. This was clearly shown in Chapter 6 Section 6.10.

8.3.2 THE SPECIFIC CONCLUSIONS TO THE RESEARCH SUB-OBJECTIVES

Sub-Objective 1: Identify the components of a good model

The main elements for highway alignment model development were identified from reviewing the existing models in the literature. It was found that a good model consists of five main parts:

- An efficient search technique
- The solution approach and the way that the problem is represented
- Formatting a testing regime (the study area)
- Defining the related cost items according to the purpose of the model
- Defining the problem constraints

Combining these items should be able to produce a realistic alignment. This was detailed in Chapter 3.

Sub-Objective 2: Identify the factors which affect alignment

The main factors which affect alignment are:

- Land use
- Land features
- Topography
- Safety and related geometric design requirements
- Cost and socio-economic activities
- Environment

These were identified by consideration of the literature in the area. This is included in Chapter 2.

Sub-Objective 3: Understand the characteristics of highway alignment geometric design

A realistic highway alignment must provide comfort and safe driving conditions. This can be achieved by adhering to the geometric design requirements. It was identified that horizontal curvature, vertical curvature, and vertical gradients are critical for the vehicles' and users' safety alongside providing comfortable driving conditions. Depending on the proposed design speed, it was shown that the minimum radius of the curve (which defines the degree of the curve or vice versa) and K value have great impacts on the horizontal and vertical curvature determinations, respectively. The details were presented in Chapter 2 Sections 2.4.1 and 2.4.2.

Sub-Objective 4: Identify the cost components that are thought critical to the problem

This study was to test the applicability of a new design approach for highway alignment development rather than generating a solution with accurate cost. Therefore, the highway related costs were reviewed and the components that were thought critical to the stated objectives were identified as:

- Length dependent costs (construction cost); to produce as short as possible alignments
- Location dependent costs; to enhance the alignment avoiding passing through high-cost fields
- Earthwork costs; to enhance the alignment avoiding high cut and/or fill areas.

These costs were identified and detailed in Chapter 2 Section 2.10.

Additional cost components can also be used for accurate cost results.

Sub-Objective 5: Understand the principles of GIS and the ways that it deals with spatial and geographic data so that a proper way is identified to format the testing regime

The spatial data relevant to the format of a study area in GIS was reviewed and the main ideas were found to be:

- Defining grids
- Defining the raster and vector data principles
- Grid cell sizes for the desired accuracy
- Using coordinated systems to deal with spatial data
- Associate the cost, land, and topographic information (e.g. cost data input) in the form of matrices with the defined grid cells. This was important to define the characteristics of different land properties and locations.

Details were given in Chapter 3 Section 3.4.2

Sub-Objective 6: Understand the elements of sustainability

A highway alignment project, for a sustainable development, should promote socio-economic generation while preserving the natural environment. Locating a highway alignment close to a land use could provide easy access and promote socio-economic generation of the area while they pay the price of proximity. Therefore, a plan for a highway should balance between the accessibility and proximity for a sustainable development. These were identified through the revision of the elements of sustainability which were included in Chapter 2 Section 2.6 and Section 2.7.

Sub-Objective 7: Investigate the possibility of formulating and solving the optimisation of a model as a genetic algorithm

The following sections describe the conclusions about the model entities by which the method was made viable to produce a ‘good’ highway alignment. Note that the words ‘simple’ or ‘complex’ for the worlds refer to the size of the world, grid cell sizes, the presence of different land features, configuration pattern of these lands, and irregularity of the topographic elevations. A small world with relatively big cells, relatively unique land features, and flat terrains was considered simple.

- Number of Station Points and Chromosome Representations

Several numbers of station points were investigated. It was found that the number of the station points was problem dependent. The irregularity (complexity) of the area, for which the alignment is proposed, has critical influence on the number of station points required to configure the alignment properly. The characteristics of the land features (irregularities), configuration pattern of these lands, and the terrain elevations were found critical. The more difficult (complex or irregular) the area, the more the station points are required to enable the alignment to skirt around the different land patterns and terrain elevations. These were proved by the experiments presented in Chapter 4 Sections 4.5.2.1 and Appendix 4. In this study, the number of the station points was related to the grid cell size and a hypothetical length of the alignment. The Equations 4.7.1 to 4.7.3 were derived to determine an approximate number of station points required to configure an alignment.

- Population Size

The results showed that the population size was problem dependent too. The more simple the world, the less the population size was required. It was found that, according to the world cases considered in the study, a population size of +5000 was important to provide the information about the whole study area. More complex worlds (in size and in land and terrain patterns) may require further investigations. One may find it difficult to comment precisely on this issue due to the strong relation of population size with the other GA elements (selection, crossover, and mutation), each of which impacts on the convergence and the quality of the solution. For instance, a small population size may lead to a premature convergence while a good selection scheme and/or crossover and mutation operator may prevent this from happening. The effect of population sizes was experimentally covered in Chapter 4 Section 4.5.2.2 and additional evaluations are presented in Appendix 5.

- Selection Scheme

Three different selection schemes, as described in Chapters 4 and 5, were investigated. All were found useful, with some discrepancies. The schemes were RnkSS, RevSS, and RndSS. RevSS outperformed RnkSS due to the incest mating characteristics of RnkSS; meanwhile, RndSS outperformed RevSS. The RnkSS and RevSS were designed as selective selection schemes while RndSS was designed to configure the solutions from information that are randomly distributed within the candidate pool. The randomness characteristics of RndSS may have promoted the chance of introducing new information to the offspring solutions.

- Crossover and Mutation Operators

It was proved that, due to the way that the alignment is represented through the X, Y, and Z coordinates of the station points, Random Multiple Point Crossover (RMPCrO) up to eight points and a mixed mutation method (GPM + SM) outperform the other tested crossover and mutation methods. A relatively long alignment passes through various lands and handles different properties and this may be the reason behind why a crossover with multiple points worked better. Also, a mutation method like GPM may help straighten some parts of the alignment and ensure different jump sizes are made if required. Details are available throughout the model formulation presented in Chapters 4 and 5.

- Individual Mutation Rate (IMR) and Point Mutation Rate (PMR)

In association with the population size of +5000, the selection of individuals to undergo mutation (IMR) was found to be 5%-10%. This range proved good to introduce new information to the solution pool and maintain the diversity of the population. On the other hand, the selection of the number of the genes (station points) per the selected individual to undergo mutation (PMR) should not exceed one gene when the Standard Mutation method (SM) is applied. It was noticed that moving more than one point could disturb and worsen the search. Moreover, 5%-15% PMR proved effective with GPM. The lower range is recommended for areas that have irregular land patterns and where the alignment requires backward and winding configurations and also where the topography of the area is also not regular; whereas, the upper range is recommended for alignments that require less winding and the elevations are more consistent.

However, it was found difficult to generalise a specific PMR range for different world cases and therefore it is recommended, as a user-defined parameter, to test various ranges even lower or higher than that specified above when a highway alignment problem is solved. This is demonstrated by the results shown in the experiment sections of Chapters 4 and 5.

8.3.3 THE DIFFICULTIES WITH THE PROPOSED METHOD

GA's operators are based on random mechanisms and therefore the direction of search is sometimes affected by the randomness of the GAs' techniques. A generated random value within the process may mislead the search and may make the solution stick at local optima. To overcome this issue, more than one test run was suggested and the research of this study was based on ten runs per test. The tuning and evaluation of the model parameters were based on the results obtained from the whole ten test runs. A search was considered stable with a set of parameters when most of the results from the ten test runs were similar and close to the optimum. By doing this, the effects of the randomness of the GA operators were reduced. This made the process of optimisation take a longer time for data analysis, decisions, and selecting the optimum solution.

8.3.4 THE MODEL EVALUATION

The model was developed based on several worlds varying between simple, specific, and relatively random and complex. The model and the search technique parameters were tuned based on the used worlds. It was found that the model is able to produce 'good' alignments when proper parameters and methods are selected. It was discovered that some specific cases might require further parameter tuning so that better solutions are obtained. Therefore, the practicality of the model could be drawn based on the produced solutions, as follows:

- It produces alignment that avoids high-cost areas
- It produces alignment that skirts around high and low elevations
- It produces vertical alignments as close as possible to the natural ground elevations
- It produces as short as possible alignments with 'relatively' straight sections where required

- It produces alignment with horizontal and vertical curvature satisfaction at any point along the alignment
- It produces alignment within the allowable maximum gradient limits
- It could produce alignments with backward characteristics
- It could handle access points (junctions) and optimise their locations along the alignment
- It could simultaneously optimise junction-highway alignment.

It should also be mentioned that the model was made flexible to respond to the planner/designer/user's different requirements. The model could produce different solutions when different cost categories and different weighting factors were used.

8.3.5 GENERAL CONCLUSIONS

The general conclusions were drawn as below.

- With the initial tests, as with many GA-based models, it was found that the whole population was dominated by identical solutions at early generations. This led to premature convergence before the entire search space was explored and even before finding the optimum solution. It was thought that using relatively larger population sizes and random-based selection schemes might have assisted in preventing early convergence occurrences. Moreover, using a crossover method like RMPCrO-8P might also have assisted in this. The conclusions drawn and presented in Section 8.3.2 give a clearer understanding.
- It was discovered that an alignment with less requirements to bends and backward bends requires higher rates of PMR than if bends and backwards are unavoidable.
- It was proved that applying repair and penalty for horizontal and vertical curvatures, as constraints, should not start at the beginning of the search. This was thought to allow a free search regardless the violation degrees of the solutions. It was shown that R&P application between the generations 25-50

could produce better results. These were experimentally shown in Chapter 6 Section 6.9.

- A number of testing regimes were designed and the model was developed accordingly. It was found difficult to design a testing regime to include every feature of real nature.
- The work described in this thesis was intended to investigate the possibility of using a new design approach. It was not intended to provide a practical tool for highway alignment design. As a research exercise the modes were tested on a wide range of artificial worlds. These were designed to simulate key aspects of real world scenarios. In future it is essential that a real world example is tested to determine if there are any significant differences between the two cases.

8.3.6 THE IMPLICATIONS OF THE MODEL IN REAL WORLD APPLICATION

The research findings address the potential of using the proposed approach of station points for alignment development in real world applications efficiently. For instance, the station points' coordinates of the resulted 3D highway alignment solution can be uploaded onto a total station whereby the alignment design can be set out on the ground without the need for IPs, tangents, curve fittings, and field calculations. The proposed approach treats the whole alignment through the coordinates of the station points. Meanwhile, the traditional methods require sequential and exhausting procedures to perform the task (refer to Chapters 1 and 2 for detail).

One may also conclude that the proposed design approach of station points can significantly reduce the time required to plan, design, and set out a highway alignment project, thus reducing the total cost of the entire process.

On the other hand, the results revealed that optimising alignments with junctions can contribute to promoting sustainable development of the area under consideration through providing efficient access while maintaining the proximity distance within the planner/designer requirements. Considering access with alignment optimisation may also raise the acceptance rate of the project by the public and the decision makers.

It is also worth mentioning that the approach can also handle judgemental factors to satisfy different criteria. This is achieved by the interpretation of the desire of the different participant parties (public, interested groups, decision makers, etc.) into different weighting factors, by which a higher or lower weight can be given to a specific desire (represented in the form of a cost function) embedded in the fitness function with which the desirability of achieving the goal could be increased or avoided. Moreover, other criteria can be satisfied and restrained through the definition of different constraint limits as presented for proximity distance satisfaction. For example, if a user considered that noise pollution was important, a factor:

$a \, . \, N$ could be included.

Where a is a multiplying factor and N is a factor representing the minimum distance required between the alignment and the specified land use.

It can also be added, as a conclusion, that the proposed approach can also be used for railways, water and gas pipelines, and corridor optimisations.

8.4 RECOMMENDATIONS FOR FUTURE WORK

The aim of the study was to investigate the possibility of improving the process of highway alignment development. Although the study has considered the most critical elements in order to test the applicability of the proposed approach for highway alignment optimisation, the work is still not complete and further investigations may still be required. The following are the recommendations for future work.

8.4.1 THE COST COMPONENTS

The fitness function of the model was formulated to test the applicability of the new proposed approach rather than producing a solution with exact cost value. Therefore, the formulation of some of the components were relatively rough and could have been refined further if producing an accurate solution was desired. Thus, for further accurate cost values the following recommendations are suggested.

8.4.1.1 Area Considerations

Square grid cells were considered for the different land areas' calculations. Although smaller grid cell sizes produce better accuracy, exact areas can be achieved by considering curved lines to represent the actual land border areas. This is achieved by either a special algorithm or by exploiting the capabilities of GIS. Using real maps representing real worlds with real geographic data are also recommended. These can be addressed in future works.

8.4.1.2 Side Slope Consideration

The land acquisition areas were considered as a function of a constant road width without taking into account the height of cut and fill and the values of side slopes. Different heights for cut and fill incur different land acquisition areas. This was also the case for earthwork calculation where no considerations were given to side slopes at different cut and fill locations. By doing this, different earthwork quantities will be produced.

8.4.1.3 Cut and Fill Sections

Cut costs in soil and rock areas are different as well as for fill costs when the availability of materials and the ground soil conditions differ between locations. A cost matrix for actual cut and fill costs for the whole area can be developed so that a more accurate calculation is obtained.

8.4.1.4 Mass Haul Costs

In addition to earthwork quantities as a cost component, hauling costs of the earthwork between the different cut and fill sections need also to be considered. Different hauling distances and the requirement for different cut and fill materials may impose different answers to the problem. This needs investigation in future works.

8.4.1.5 User Costs

This issue by itself is an independent research subject. The implications and consequences of user costs are not explicit in the literature. User costs were mainly considered in terms of fuel consumption, travel time, and accident costs. Many other factors may also affect this cost as described in Chapter 2 Section 2.10.1.4. The effect of these elements on alignment length, environment, the alignment geometric design, safety, and design speed may need an extensive coverage.

8.4.1.6 Important Road Structures

The cost of some important road structures may have great influence on the whole process. The inclusion of structures like bridges, interchanges, and tunnels may produce different answers. This could be addressed independently and extensively.

8.4.1.7 Superelevation and Widening at Bends

Superelevation and widening are two elements of the traditional geometric design element which are considered for safer driving conditions at curves. This issue needs to be incorporated with the proposed design approach for future investigation in two directions: first as elements of cost which affect the quantities of earthwork, construction material, and the areas covered by the alignment; and second, as elements of geometric design for safety.

8.4.2 GA OPERATORS

As the model was built under the light of a new design approach, only several GA operators were investigated. With these operators it was proved that the approach is valid and assists producing an optimum solution within a relatively good time. Station point approach, as a new design technique, requires further investigation using other combinations of GA operators. For instance, using a mutation method for Z coordinates that works on a station point (as a gene) within a specified limit of ground elevations (e.g. within ±10m cut or fill depths of the station point's ground elevation).

8.4.3 JUNCTION RELATED ISSUES

8.4.3.1 The Junction-Highway Alignment Optimisation

In this study, the optimum junction locations and their effects on the main alignment's location and configuration were considered. The experiments were carried out on 2D highway alignments. Moreover, the access links between the land use centres and the junctions on the main alignment were considered straight and direct (instantaneous). Further studies are required to deal with access provision for 3D highway alignment and to take into account the real land properties between the land use centres and the junction so that the link path is not straight. This is expected to realistically address the effect of access provision on the main highway alignment.

8.4.3.2 Junction Costs and Safety Consideration

Optimum junction locations are not only associated with the land use centre's locations. The amount of traffic from two or more land use centres may suggest combining two or more junctions in one if they require access points close to each other. The type, size, and cost of the junctions need also to be investigated. Moreover, the distances between the successive junctions may have effects on safety and traffic disturbances. Junction safety consideration should try to locate the junctions at positions that ensure smooth and continuous flow with less traffic interruptions. These are factors that require further research.

8.4.3.3 Public Participation and Decision Making

Further studies can be made so that the model allows different input values from public and other interested groups. Different land use centres may have different inputs into the system. These values may give different weights to the requirement for access and proximity. The corresponding junction locations with respect to the different land use requirements, as input, need to be investigated whereby the location and configuration of the main alignment is found.

8.5 SUMMARY

This chapter has summarised the main findings and put forward the outlines for future studies. The study was successful in achieving its aims to optimise highway alignments through station points as a new design method. It was also suggested that junctions can also be handled by the proposed design approach and recommends simultaneous junction-highway alignment optimisation.

The success of introducing the notion of station points for highway alignment development, including the inclusion of junctions, opened a broad argument for further investigations.

REFERENCES

Allan, A.L, (2007). *Principles of Geospatial Surveying*, Whittles Publishing, UK.

American Association of State Highway and Transportation Officials, (1994). USA.

Alterkawi, M., (2001). Application of GIS in Transportation Planning: The Case of Riyadh, the Kingdom of Saudi Arabia, *GBER*, Vol. 1, No. 2, pp. 38-46.

Andre, J., Siarry, P. & Dognon, T., (2001). An Improvement of the Standard genetic Algorithm Fighting Premature Convergence in Continuous Optimisation, *Advances in* Engineering Software, Vol. 32, pp. 49-60.

Ashworth, R., (1966). *Highway Engineering*, Heinemann Educational Books Ltd., UK.

Baker, J. E., (1985). Adaptive Selection Methods for Genetic Algorithms. *Proceeding of the 3rd International Conference on Genetic Algorithms and Applications*, New Jersey, Lawrence Erlbaum: Hillsdale, pp. 100-111.

Beasley, D., Bull, D. R. & Martin, R. R., (1993). An Overview of Genetic Algorithms: Part 1 Fundamentals, *University Computing*, Vol. 15, Issue 2, pp. 58-69.

Beasley, D., Bull, D. R. & Martin, R. R., (1993). An Overview of Genetic Algorithms: Part 2, Research Topics, *University Computing*, Vol. 15, Issue 4, pp. 170-181.

Blickle, T. & Thiele, L., (1995). A Comparison of Selection Schemes Used in Genetic Algorithms, Swiss Federal Institute of Technology (ETH) Zurich, *Computer Engineering and Communications Networks Lab, TIK Report*, Volume 2, No. 11, Swiss.

Brundtland, G. H., (1987). *Our Common Future: Report of the World on Environment and Development*, Centre of a World in Balance, Netherlands. Available from <http://www.-worldinbalance.net/agreements/1987-brundtland.php>. [Accessed: 03 January 2009].

Bruton, M. J., (1992). *Introduction to Transport Planning*, UCL press, UK.

Bussieck, M.R. & Pruessner, R., (2003). *Mixed-Integer Nonlinear Programming*, GAMS Development Corporation, Washington, USA, available from <http://www.gamsworld.org/-minlp/siagopt.pdf>. [Accessed: 28 June 2011].

Chakraborty, R.C., (2010). *Fundamental of Genetic Algorithms*, Artificial Intelligence, Defence Research & Development Organisation (DRDO), India. Available from <http://-www.myreaders.info/html/artificial_intelligence.html>. [Accessed: 18 June 2011]

Cheng, J. & Lee, Y., (2006). Model for Three Dimensional Highway Alignment, *Journal of Transportation Engineering*, Vol. 132, No. 12, pp. 913-920.

Chew, E.P., Goh, C.J. & Fwa, T.F., (1989). Simultaneous Optimization of Horizontal and Vertical Alignments of Highways, *Transportation research*, Vol. 23B, No. 5, pp. 315-329.

Coley, D. A., (1999). *An Introduction to Genetic Algorithms for Scientists and Engineers*, World Scientific Publishing Co. Pte.Ltd., Singapore.

Cowen, D. J., (1988). GIS Versus CAD Versus DBMS: What are the Differences?, *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 11, pp. 1551-1554.

Davis, L., (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.

Design Manual for Roads and Bridges, (1992-2008). The Highways Agency, UK.

de Smith, M. J., (2006). Determination of Gradient and Curvature Constrained Optimal Paths, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 21, Issue 1, pp. 24-38.

Delavar, M. R., Samadzadegan, F. & Pahlavani, P., (2004). A GIS Assisted Optimal Urban Route Finding Approach Based on genetic Algorithms, *XXth ISPRS Congress: Geo-Imagery Bridging Continents*, Istanbul, Turkey.

Demers, M. N., (2005). *Fundamentals of Geographic Information System*, John Wiley and Sons, Inc., USA.

Easa, S. M., Strauss, T. R., Hassan, Y. & Souleyrette, R., (2002). Three-Dimensional Transportation Analysis: Planning and Design, *Journal of Transportation Engineering*, Vol. 128, No. 3, pp. 250-258.

Easa, S.M., (1988). Selection of Roadway Grades that Minimise Earthwork Cost Using Linear Programming, *Transportation Research*, Vol. 22A, No. 2, pp. 121-136.

Easa, S. M. & Mehmood, A., (2008). Optimizing Design of Highway Horizontal Alignments: New Substantive Safety Approach, *Computer-Aided Civil and Infrastructure Engineering*, 23, pp 560 – 573.

Eshelman, L. J. & Schaffer, J. D., (1991). Preventing Premature Convergence in Genetic Algorithms by Preventing Incest, *International Conference on Genetic Algorithms - ICGA*, pp. 115-122.

Federal Highway Administration (FHWA), Flexibility in Highway Design, (1997), U.S. Department of Transportation, USA. Available on < http://ttap.colostate.edu/Library-/FHWA/FHWA-PD-97-062.pdf>. [Accessed: 05 December 2008]

Ford, M., (2007). *A Genetic Algorithm Based Decision Support System for the Sustainable Location of development Allocations*, Ph.D. Thesis, University of Nottingham, University Park.

Forrest, S., (1996). Genetic Algorithms, *ACM Computer Surveys*, Vol. 28, No. 1, pp. 77-80.

Fwa, T. F., Chan, W. T. & Sim, Y. P., (2002). Optimal Vertical Alignment Analysis for Highway Design, *Journal of Transportation Engineering,* Vol. 128, No. 5, pp. 395 – 402.

Fwa, T.F., (1989). Highway Vertical Alignment Analysis by Dynamic Programming, *Transportation Research Record*, Vol. 1239, pp. 1-9.

Galati, S. R., (2006). *Geographic Information Systems Demystified*, Artech house INC., UK.

Garber, N. J. & Hoel, L. A., (2009). *Traffic and Highway Engineering*, Cengage Learning by Nelson Education Ltd., USA.

Goh, C.J., Chew, E.P. & Fwa, T.F., (1988). Discrete and Continuous Models for Computation of Optimal Vertical Highway Alignment, *Transportation Research*, Vol. 22B, No. 6, pp. 399-409.

Goktepe, A. B., Lav, A. H. & Altun, S., (2005). Dynamic Optimization Algorithms for Vertical Alignment of Highways, *Mathematical and Computational Applications*, Vol. 10, No. 3, pp. 341 -350.

Goktepe, A. B. & Lav, A. H., (2003). Method for Balancing Cut-Fill and Minimizing the Amount of Earthwork in the Geometric Design of Highways, *Journal of Transportation Engineering*, Vol. 129, No. 5, pp. 564 - 571.

Goktepe, A.B., Altun, S. & Ahmedzade, P., (2009). Optimization of Vertical Alignment of Highways Utilizing Discrete Dynamic Programming and Weighted Ground Line, *Turkish Journal of Engineering and Environmental Sciences*, Vol. 33, Issue 2, pp. 105-116.

Goldberg, D.E. & Deb, K., (1991). A Comparative analysis of Selection Schemes used in Genetic Algorithms, *In Foundations of Genetic Algorithms*, Rawlins, G.J.E., Morgan Kaufmann Publishers, USA, pp. 69-93.

Haq, G., (1997). *Towards Sustainable Transport Planning*, Ashgate publishing limited, England.

Hassan, Y., Easa, S.M. & ABD EL Halim, A.O., (1998). Highway Alignment: Three-Dimensional Problem and Three-Dimensional Solution, *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1612, pp. 17-25.

Haupt, R. L. & Haupt, S. E., (2004). *Practical Genetic Algorithms*, John Wiley & Sons, Inc, USA.

Hayman, R.W., (1970). Optimisation of Vertical Alignment for Highways through Mathematical Programming, *Highway Research Record*, No. 306, Highway Research Board, pp. 1-9.

Howard, B.E., Bramnick, Z. & Shaw, J.F.B, (1969). Optimum Curvature Principle in Highway Routing, *Journal of the Highway Division*, ASCE, Vol. 94, No. HW1, pp. 61-82.

Heywood, I., Carnelius, S. & Carver, S., (2006). *An Introduction to Geographic Information System*, Person education limited, England.

Hrstka, O. & Kucerova, A., (2004). Improvements of Real Coded Genetic Algorithms Based on Differential Operators Preventing Premature Convergence, *Advances in Engineering Software*, Vol. 35, pp. 237-246.

Jaarsma, C. F., (1997). Approaches for the planning of rural road network according to Sustainable land use planning, *Landscape and Urban Planning*, Vol. 39, pp. 47-54.

Jha, M.K., (2002). Optimising Highway Networks: A genetic Algorithms and Swarm Intelligence Based Approach, *Proceedings of International Workshop on Information Technology in Civil Engineering*, Washington, USA.

Jha, M. K. & Kim, E., (2006). Highway optimization based on accessibility, proximity and land use change, *Journal of Transportation Engineering*, Vol. 132, No.5, pp. 435-439.

Jha, M. K. & Maji, A., (2007). A Multi-Objective Genetic Algorithm for Optimizing Highway Alignments, *The IEEE Symposium on Computational Intelligence in Multi-criteria Decision Making (MCDM)*, Honolulu, HI, USA.

Jha, M. K., McCall, C. & Schonfeld, P., (2001). Using GIS, Genetic Algorithms, and Visualization in Highway Development, *Compute-Aided Civil and Infrastructure Engineering*, Blackwell Publishers, Vol. 16, No. 6, pp. 399-414, USA.

Jha, M. K. & Schonfeld, P., (2000). Integrating Genetic Algorithms and Geographic Information System to Optimize Highway Alignments, *Transportation Research Record*, Vol. 1719, pp. 233-240.

Jha, M. K., Schonfeld, P., Jong, J. C. & Kim, E., (2006). Intelligent Road Design, WITpress, UK.

Jha, M. K., (2003). Criteria-Based Decision Support System for Selecting Highway Alignments, *Journal of Transportation Engineering*, Vol. 129, No.1, pp. 33-41.

Jong, J. C., (1998). *Optimizing Highway Alignments with Genetic Algorithms*, Ph.D. dissertation, University of Maryland, College Park, USA.

Jong, J. C., Jha, M. K., McCall, C. & Schonfeld, P., (2000). Preliminary highway Design with Genetic Algorithms and Geographic Information Systems, *Computer-Aided Civil and Infrastructure Engineering*, Vol.15, Issue 4, pp. 261-271.

Jong, J. C. & Schonfeld, P., (1999). Cost Functions for Optimizing Highway Alignments, *Transportation Research Record*, Vol. 1659, pp. 58-67.

Jong, J.C. & Schonfeld, P., (2003). An evolutionary model for simultaneously optimizing three-dimensional highway alignments, *Transportation Research*, Part B, Volume 37, Issue 2, pp. 107-128.

Kalamaras, G.S., Brino, L., Carrieri, G., Pline, C. & Grasso, P., (2000). Application of Multicriteria Analysis to Select the Best Highway Alignment, *Tunneling and Underground Space Technology*, Vol. 15, No. 4, pp. 415-420.

Kang, M.W., (2008). *An Alignment Optimisation Model for a Simple Highway Network*, PhD Thesis, University of Maryland, College Park, USA.

Kang, M. W., Schonfeld, P. & Jong, J. C., (2007). Highway Alignment Optimization through Feasible Gates, *Journal of Advanced Transportation*, Vol. 41, No.2, pp. 115-144.

Kang, M.W., Schonfeld, P. & Yang, N., (2009). Prescreening and Repairing in a Genetic Algorithm for Highway Alignment Optimization, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 24, pp. 109-119.

Kersten, G.E., Mikolajuk, Z. & Yeh, A.G., (1999). *Decision Support system for Sustainable Development*, A research Book of Methods and Applications, Kluwer Academic Publishers, Hingham, MA, USA.

Khatib, Z., Ou, Y., Chang, K., (1999), *GIS and Transportation Planning, Research and Innovative Technology Administration (RITA)*, U.S. Department of Transportation (US DOT), USA, Sixth National Conference on Transportation Planning for Small and Medium-sized Communities. Available on < http://ntl.bts.gov/lib/000/500/570/00780106.-pdf>. [Accessed: 14 November 2008].

Kim, E., Jha, M.K. & Son, B., (2005). Improving the Computational Efficiency of Highway Alignment Optimization Models through a Stepwise Genetic Algorithms Approach, *Transportation Research Part B*, Vol. 39, pp. 339-360.

Kim, E., Son, B., Kwon, Y. & Lee, H., (2003). Incorporating Simple Grade Separation Structures into Highway Alignment Optimisation, *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 5, pp. 884-897.

Kitamura, R. & Fujii, S., (1997). Time Use Data, Analysis and Modeling: Toward the Next Generation of Transportation Planning Methodologies, *Transport Policy*, Vol. 4, No. 4, pp. 225-235.

Krammes, R.A & Granham, M.A., (1995). Worldwide Review of Alignment Design Principles, *Proceedings of International Symposium on Highway Geometric Design Practices*, Boston, USA.

Lee, Y., Tsou, Y.R. & Liu, H.L., (2009). Optimization Method for Highway Horizontal Alignment Design, *Journal of Transportation Engineering*, Vol. 135, No. 4, pp. 217-224.

Lee, Y. & Cheng, J., (2001). Optimising Highway Grades to Minimize Cost and Maintain Traffic Speed, *Journal of Transportation Engineering*, Vol. 127, No. 4, pp. 303-310.

Lee, C.Y., (2003). Entropy-Boltzmann selection in the genetic algorithms. *IEEE Transactions on Systems*, Man, and Cybernetics, Part B, Vol. 33, No. 1. pp 33, 138-142.

Longley, P.A., Goodchild, M.F., Maguire, D.J. & Rhind, D.W., (2001). *Geographic Information Systems and Science*, John Wiley and Sons Ltd, UK.

Maguire, D.J., Batty, M. & Goodchild, M.F., (2005). *GIS, Spatial Analysis, and Modeling*, ESRI Press, USA.

Maji, A. & Jha, M.K., (2009). Multi-objective highway alignment optimization using a genetic algorithm, *Journal of Advanced Transportation*, Vol. 43, Issue 4, PP. 481-504.

Mannering, F.L., Washburn, S.S. & Kilareski, W.P., (2009). *Principles of Highway Engineering and Traffic Analysis*, John Wiley & Sons Inc., USA.

Matthews, K.B., Craw, S., MacKenzie, I., Elder, S. & Sibbald, A. R., (1999). Applying Genetic Algorithms to Land Use Planning, *Proceedings of the 18th Workshop of the UK Planning and Scheduling Spatial Interest Group*, University of Salford, PP 109-115, 15th – 16th Dec.

Mawdesley, M.J., Al-Jibouri, S.H. & Yang, H., (2002). Genetic Algorithms for Construction Site Layout in Project Planning, *Journal of Construction Engineering and Management*, Vol (128), No. 5, pp. 418-426.

Meyer, M.D. & Miller, E.J., (2001). *Urban Transportation Planning*, McGraw-Hill.

Michalewicz, Z., (1999). *Genetic Algorithms + Data Structures = Evolution Programs*, Third Revised and Extended Edition, Springer-Verlag, New York.

Mitchell, M., (1996). *An Introduction to Genetic Algorithms*, Massachusetts Institute of Technology (The MIT Press), USA.

Moreb, A. A., (1996). Linear Programming Model for Finding Optimal Roadway Grades that Minimize Earthwork Cost, *European Journal of Operational Research*, Vol. 93, pp. 148 – 154.

Moreb, A. A., (2009). Spline Technique for Modeling Roadway Profile to Minimize Earthwork Cost, *Journal of Industrial and Management Optimization*, Vol. 5, No. 2, pp. 275 – 283.

Murray, T. Alan., (2001). Strategic Analysis of Public Transport Coverage, *Socio-Economic Planning Sciences*, Vol. 35, pp. 175-188.

O'Flaherty, C.A., (1991). *Highways*, Edward Arnold, Volume 2, UK.

Olver, P.J., (2010). *Calculus of Variations*, University of Minnesota, College of Science and Engineering, USA. Available on <http://www.math.umn.edu/~olver/-am_/cvz.pdf>. [Accessed: 28 June 2011].

Parker, N.A., (1977). Rural Highway Route Corridor Selection, *Transportation Planning and Technology*, Vol. 3, pp. 247-256.

Piantanakulchai, M., (2005). Analytic Network Process Model for Highway Corridor Planning, *ISAHP*, Honolulu, Hawaii, July 8-10, USA.

Rhind, D.W., (1988). A GIS Research Agenda, *International Journal of Geographic Information Systems*, Vol 2, pp. 23-28.

Roberts, P.O., (1957). Using New Methods in Highway location, *Photogrammetric Engineering*, Vol. 23, No. 3, pp. 563-569.

Rogers, M., (2003). *Highway Engineering*, Blackwell Publishing, UK.

Saffarzadeh, M. & Broujerdian, A.M., (2007). Optimum Route Location Model Considering the Trip Generation Centers and Protected Zones, *Iranian Journal of Science and Technology (Transaction B, Engineering)*, Vol. 31, No.B2, pp. 167-178.

Salter, R.J. and Hounsell, N.B., (1996). *Highway Traffic Analysis and Design*, Palgrave Macmillan, UK.

Schofield, W. & Breach, M., (2007). *Engineering Surveying*, Elsevier Ltd, UK.

Schoon, J.G., (2001). *Geometric Design Projects for Highways: an Introduction*, ASCE Publication, Reston, VA, USA.

Shaw, J.F.B. and Howard, B.E., (1982). Expressway Route Optimization by OCP, *Journal of Transportation Engineering*, ASCE, 108(3), 227-243.

Smith, D. K., (1991). *Dynamic Programming: A Practical Introduction*, Ellis Horwood Limitted, England.

Social Science Network and Data Services (ssnds), (2006), *Real World Representation in GIS*, The University of Western Ontario, Ontario, Canada.

Available on <http://ssnds.-uwo.ca/sscnetworkupdate/2006winter/images/-gis.jpg>. [Accessed: 12 November 2008].

Sthapit, N. & Mori, H., (1994). Model to estimate Highway Earthwork Cost in Nepal, *Journal of Transportation Engineering*, Vol. 120, Issue 3, pp. 498-504.

Tang, K.X. & Waters, N.M., (2005). The Internet, GIS and Public Participation in Transportation Planning, *Progress in Planning*, Vol. 64, Issue 1, pp. 7-62.

Tat, C.W. & Tao, F., (2003). Using GIS and Genetic Algorithms in Highway Alignment Optimization, *Intelligent Transportation System*, Vol.2, pp. 1563-1567.

Uren, J. & Price, B., (2010). *Surveying for Engineers*, Palgrave Macmillan, UK.

Walsh, G. R., (1985). *An Introduction to Linear Programming*, John Wiley and Sons, England.

Wan, F.Y.M., (1995). *Introduction to the Calculus of Variations and its Applications*, Chapman & Hall, New York, USA.

Wikipedia, (2011). Calculus of Variations, Available on <http://en.wikipedia.org/wiki/-Calculus_of_variations>. [Accessed: 15th March 2011]

Woods, K.B., Berry, D.S. & Goetz, W.H., (1960). *Highway Engineering Handbook*, McGraw Hill, USA.

Wright, P.H. & Ashford, N.J., (1998). *Transportation Engineering: Planning and Design*, Fourth Edition, John Wiley and Sons, USA.

Yang, Z., Feng, T. & Peng, J., (2003). Optimising Highway Alignment in Road Network, *Proceeding of the Eastern Asia Society for Transportation Studies*, Vol. 4, pp. 373-380.

Zegeer, C.V, Stewart, J.R., Council, F.M., Reinfurt, D.W., and Hamilton, E., (1992). Safety Effects of Geometric Improvements on Horizontal Curves, *Transportation Research Board*, No 1356, pp. 11-19.

Zhang, X. and Armstrong, M.P., (2008). Genetic algorithms and the corridor location problem: multiple objectives and alternative solutions, *Environment and Planning B: Planning and Design*, Pion Ltd, London, vol. 35, Issue (1), pp. 148-168.

APPENDIX 1

DESIGN VALUES OF K FACTOR FOR VERTICAL CURVES

The design values of K for Crest and Sag vertical curves, according to the specified design speeds, are used to define proper vertical curve lengths. The resulted curve lengths will satisfy both the required curvature for smooth and safe driving condition, and the length required for Stopping Sight Distance (SSD).

Table APP1.1: Design values of K for CREST vertical curves based on SSD

U.S. Customary				Metric			
Design speed (mi/hr)	SSD (ft)	Rate of vertical curvature, K*		Design speed (km/hr)	SSD (m)	Rate of vertical curvature, K*	
		Calculated	Design			Calculated	Design
15	80	3	3	20	20	0.6	1
20	115	6.1	7	30	35	1.9	2
25	155	11.1	12	40	50	3.8	4
30	200	18.5	19	50	65	6.4	7
35	250	29	29	60	85	11.0	11
40	305	43.1	44	70	105	16.8	17
45	360	60.1	61	80	130	25.7	26
50	425	83.7	84	90	160	38.9	39
55	495	113.5	114	100	185	52.0	52
60	570	150.6	151	110	220	73.6	74
65	645	192.8	193	120	250	95.0	95
70	730	246.9	247	130	285	123.4	124
75	820	311.6	312				
80	910	383.7	384				

*Source: (AASHTO, 2004) (Mannering et al, 2009) (Garber and Hoel, 2009)

Table APP1.2: Design values of K for SAG vertical curves based on SSD

U.S. Customary				Metric			
Design speed (mi/hr)	SSD (ft)	Rate of vertical curvature, K*		Design speed (km/hr)	SSD (m)	Rate of vertical curvature, K*	
		Calculated	Design			Calculated	Design
15	80	9.4	10	20	20	2.1	3
20	115	16.5	17	30	35	5.1	6
25	155	25.5	26	40	50	8.5	9
30	200	36.4	37	50	65	12.2	13
35	250	49	49	60	85	17.3	18
40	305	63.4	64	70	105	22.6	23
45	360	78.1	79	80	130	29.4	30
50	425	95.7	96	90	160	37.6	38
55	495	114.9	115	100	185	44.6	45
60	570	135.7	136	110	220	54.4	55
65	645	156.5	157	120	250	62.8	63
70	730	180.3	181	130	285	72.7	73
75	820	205.6	206				
80	910	231	231				

*Source: (AASHTO, 2004) (Mannering et al, 2009) (Garber and Hoel, 2009)

APPENDIX 2

THE MODEL INTERFACE FORM WINDOWS

The following presentations show the model interfaces form windows with the most input options. They also illustrate how a test run was prepared using the developed model interface options. The parameters for a realistic 3D highway alignment are shown within the input boxes of the model interface. The selected options consider Horizontal curvature, vertical curvature, and vertical line gradients as constraints. The results of the associated parameters are also presented.

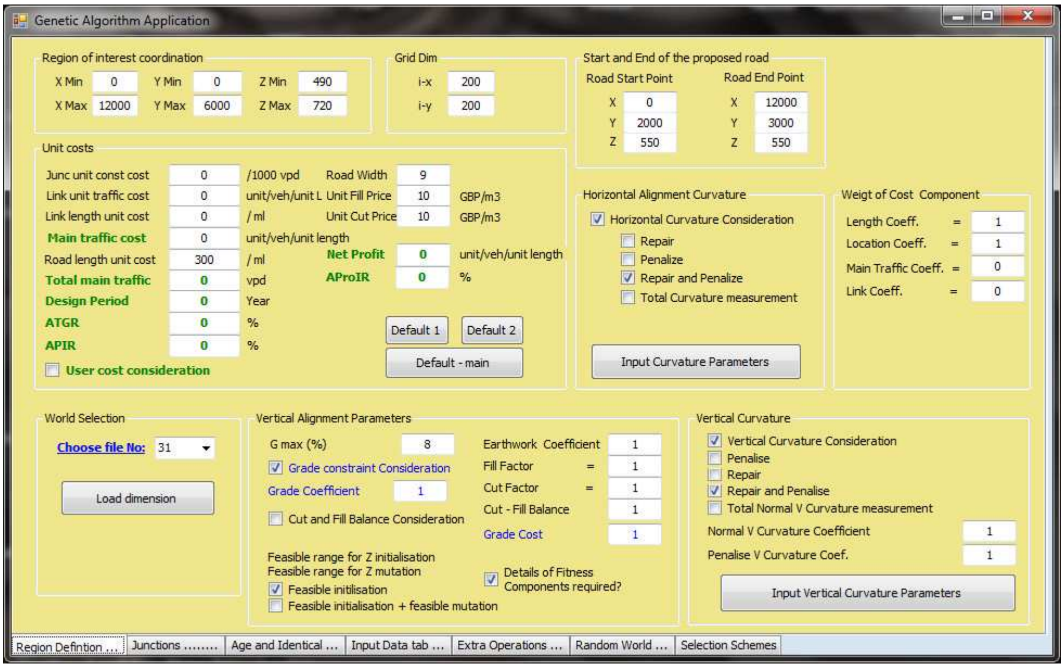


Figure APP2.1: The data input main interface

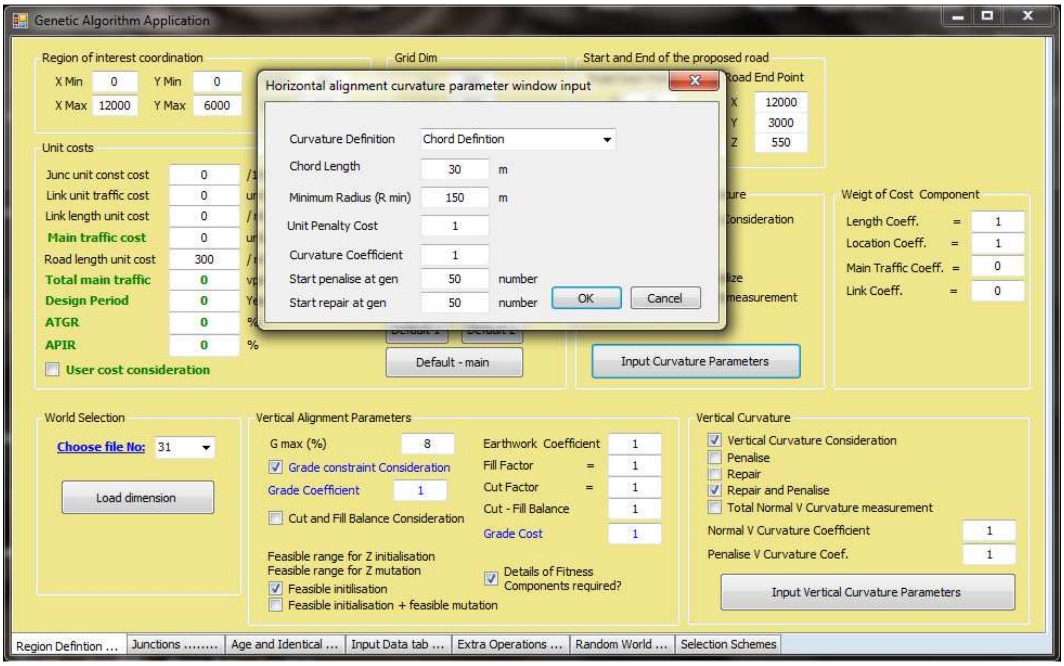


Figure APP2.2: The data input window for horizontal curvature within the main interface

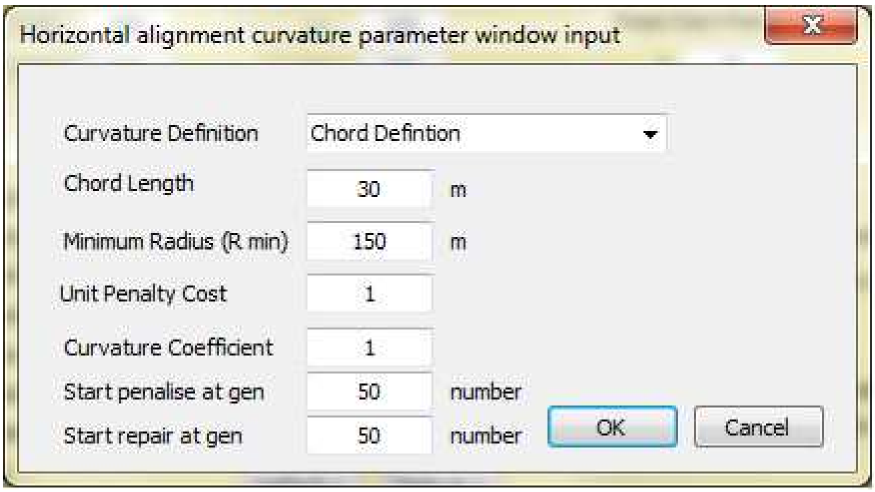


Figure APP2.3: The data input window for horizontal curvature parameters

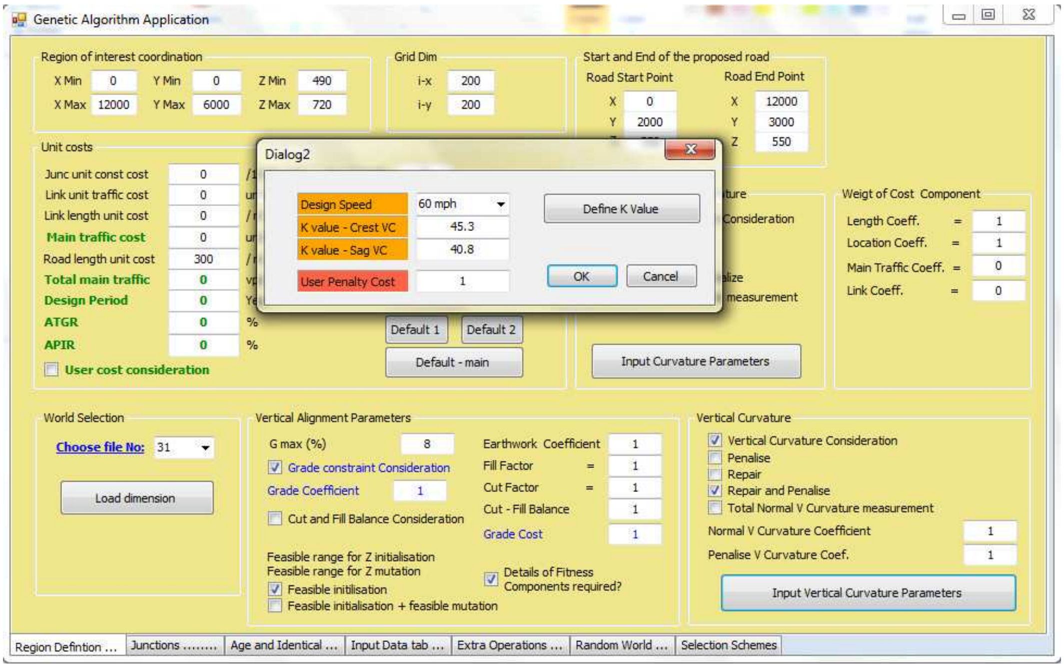


Figure APP2.4: The data input window for vertical curvature within the main interface

Figure APP2.5: The data input window for vertical curvature parameters

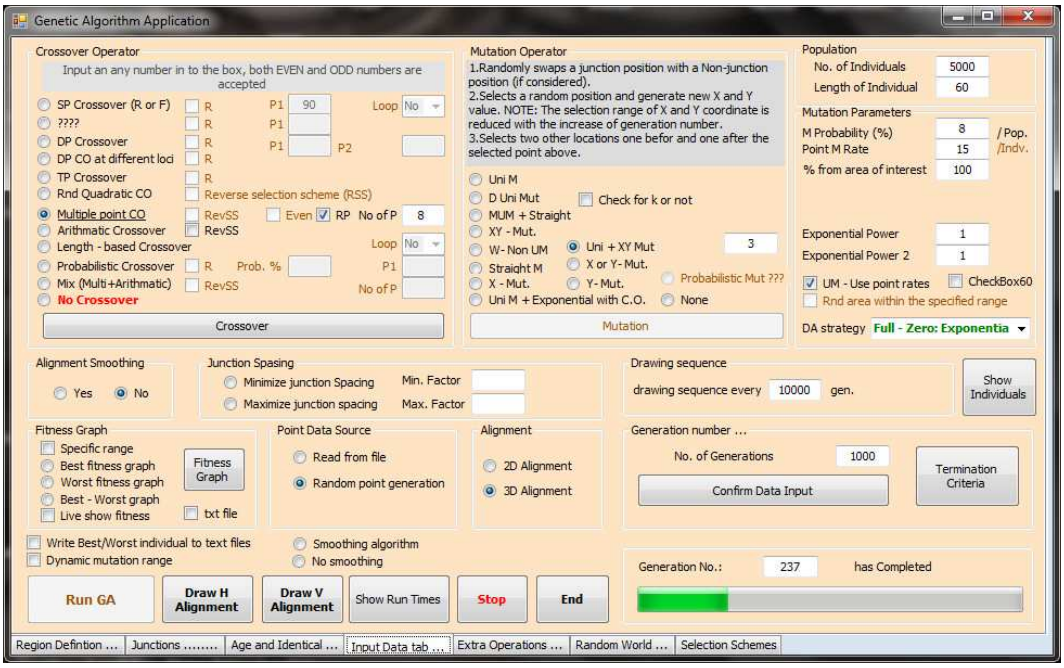


Figure APP2.6: The main interface for the GA parameter settings

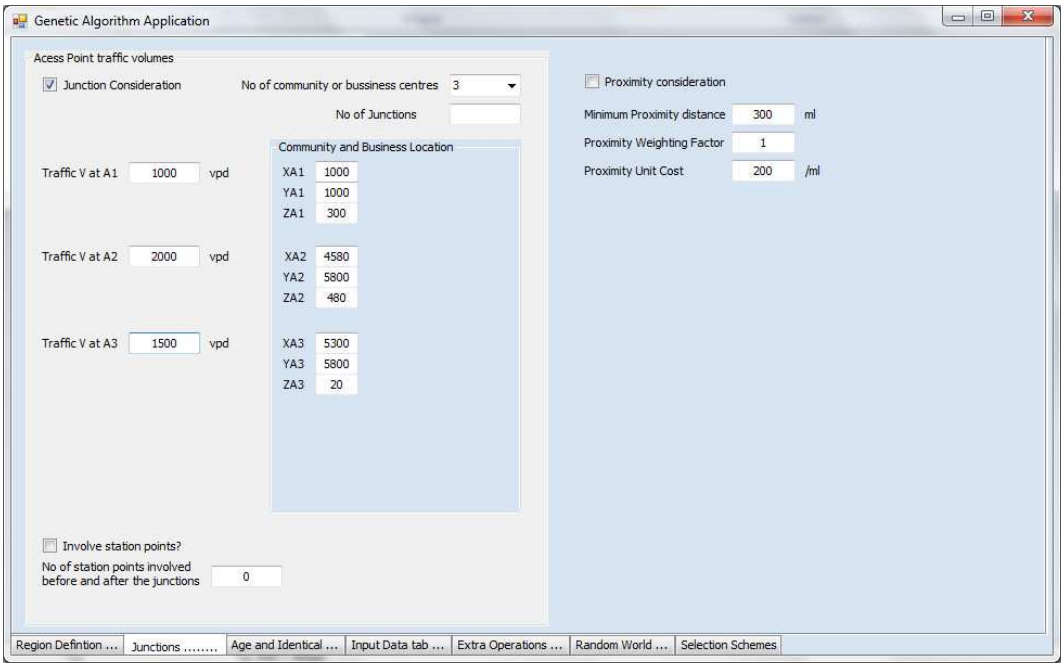


Figure APP2.7: The main input interface for junction consideration

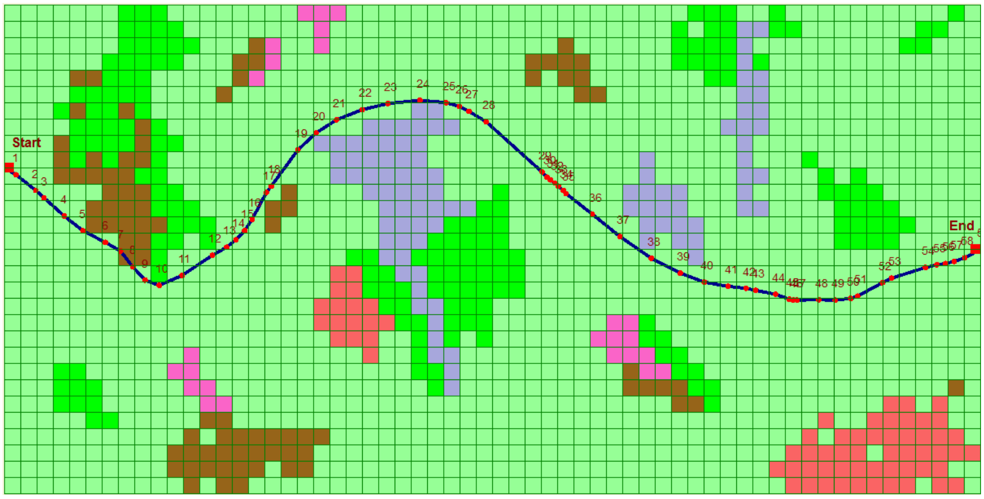


Figure APP2.8: A highway alignment at gen 1000 with station point numbers (land feature map)

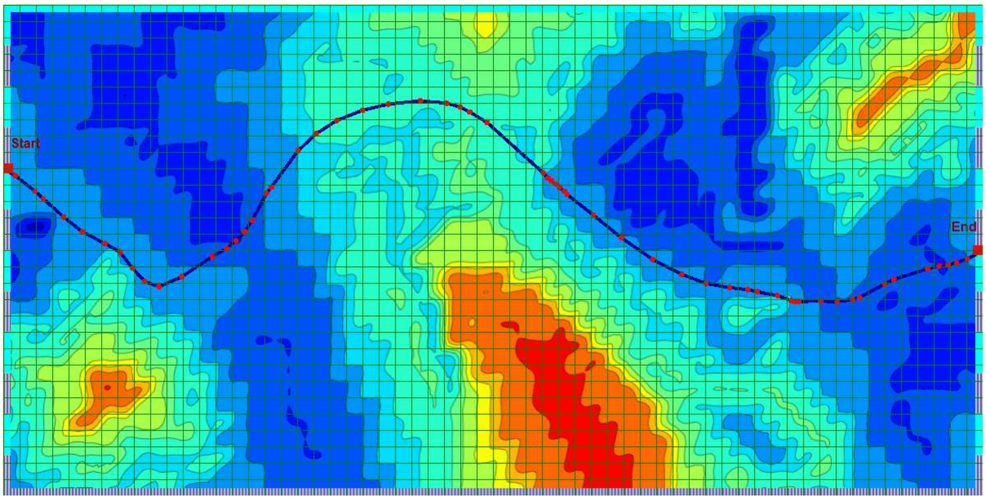


Figure APP2.9: A highway alignment at generation 1000 (contour map)

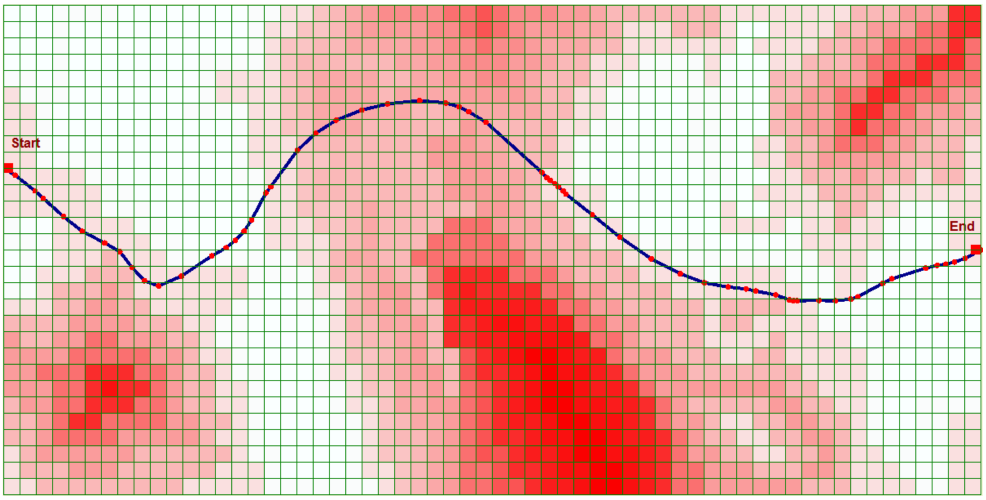
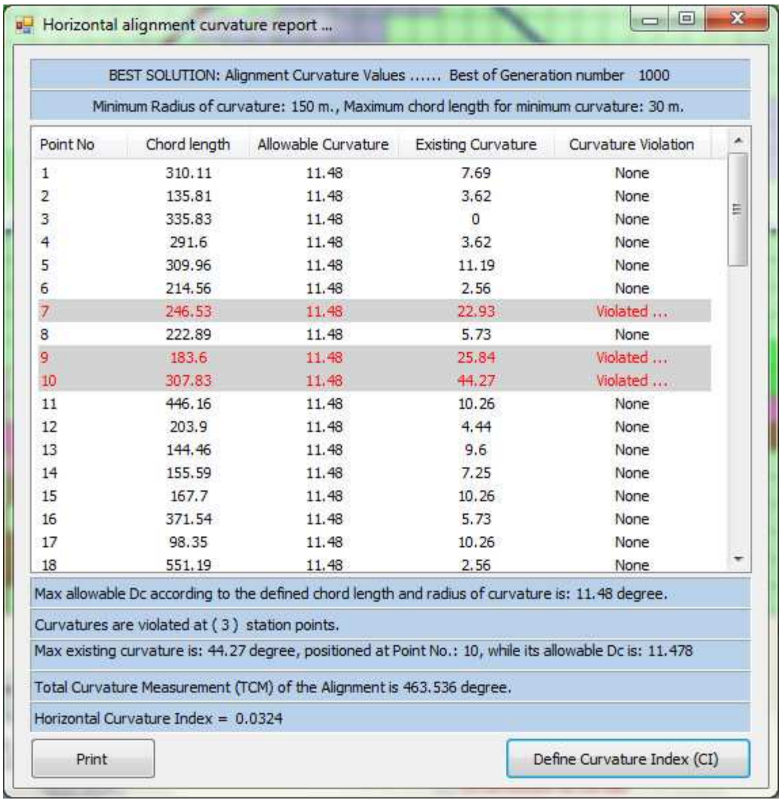


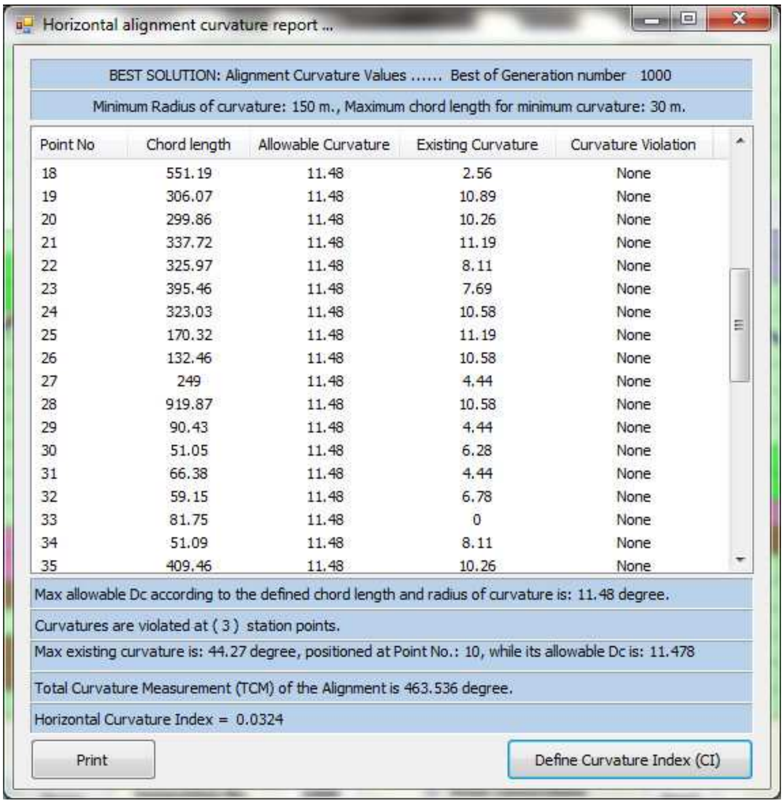
Figure APP2.10: A highway alignment at generation 1000 (digital elevation map)

The Figures shown below are the curvature reports for both horizontal and vertical alignments including the gradient values between the successive station points.



Point No	Chord length	Allowable Curvature	Existing Curvature	Curvature Violation
1	310.11	11.48	7.69	None
2	135.81	11.48	3.62	None
3	335.83	11.48	0	None
4	291.6	11.48	3.62	None
5	309.96	11.48	11.19	None
6	214.56	11.48	2.56	None
7	246.53	11.48	22.93	Violated ...
8	222.89	11.48	5.73	None
9	183.6	11.48	25.84	Violated ...
10	307.83	11.48	44.27	Violated ...
11	446.16	11.48	10.26	None
12	203.9	11.48	4.44	None
13	144.46	11.48	9.6	None
14	155.59	11.48	7.25	None
15	167.7	11.48	10.26	None
16	371.54	11.48	5.73	None
17	98.35	11.48	10.26	None
18	551.19	11.48	2.56	None

Figure APP2.11: Horizontal curvature report output window (part 1)



Horizontal alignment curvature report ...

BEST SOLUTION: Alignment Curvature Values ...... Best of Generation number 1000

Minimum Radius of curvature: 150 m., Maximum chord length for minimum curvature: 30 m.

Point No	Chord length	Allowable Curvature	Existing Curvature	Curvature Violation
18	551.19	11.48	2.56	None
19	306.07	11.48	10.89	None
20	299.86	11.48	10.26	None
21	337.72	11.48	11.19	None
22	325.97	11.48	8.11	None
23	395.46	11.48	7.69	None
24	323.03	11.48	10.58	None
25	170.32	11.48	11.19	None
26	132.46	11.48	10.58	None
27	249	11.48	4.44	None
28	919.87	11.48	10.58	None
29	90.43	11.48	4.44	None
30	51.05	11.48	6.28	None
31	66.38	11.48	4.44	None
32	59.15	11.48	6.78	None
33	81.75	11.48	0	None
34	51.09	11.48	8.11	None
35	409.46	11.48	10.26	None

Max allowable Dc according to the defined chord length and radius of curvature is: 11.48 degree.

Curvatures are violated at (3) station points.

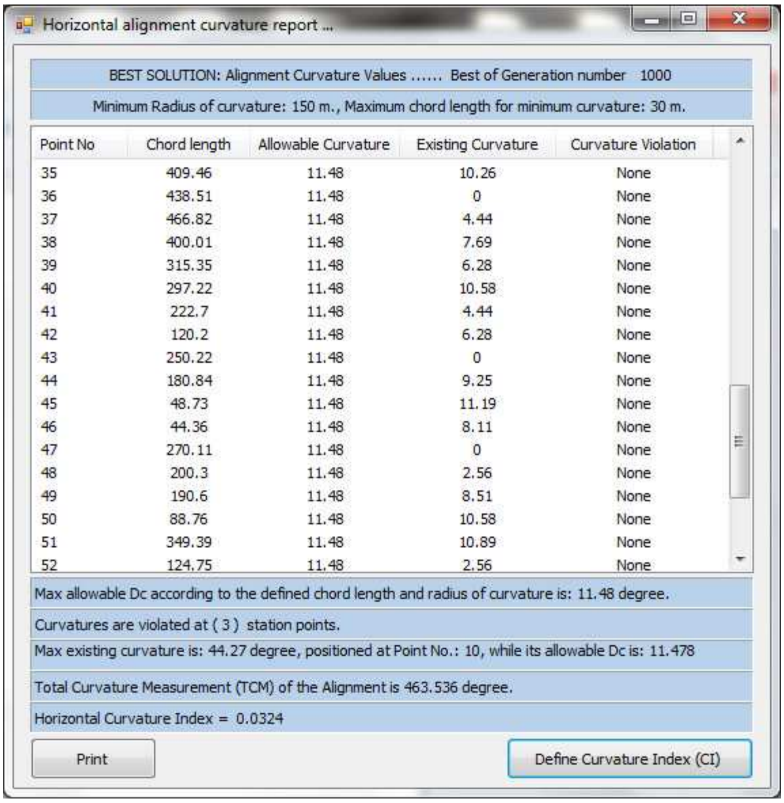
Max existing curvature is: 44.27 degree, positioned at Point No.: 10, while its allowable Dc is: 11.478

Total Curvature Measurement (TCM) of the Alignment is 463.536 degree.

Horizontal Curvature Index = 0.0324

Print | Define Curvature Index (CI)

Figure APP2.12: Horizontal curvature report output window (part 2)



Horizontal alignment curvature report ...

BEST SOLUTION: Alignment Curvature Values ...... Best of Generation number 1000

Minimum Radius of curvature: 150 m., Maximum chord length for minimum curvature: 30 m.

Point No	Chord length	Allowable Curvature	Existing Curvature	Curvature Violation
35	409.46	11.48	10.26	None
36	438.51	11.48	0	None
37	466.82	11.48	4.44	None
38	400.01	11.48	7.69	None
39	315.35	11.48	6.28	None
40	297.22	11.48	10.58	None
41	222.7	11.48	4.44	None
42	120.2	11.48	6.28	None
43	250.22	11.48	0	None
44	180.84	11.48	9.25	None
45	48.73	11.48	11.19	None
46	44.36	11.48	8.11	None
47	270.11	11.48	0	None
48	200.3	11.48	2.56	None
49	190.6	11.48	8.51	None
50	88.76	11.48	10.58	None
51	349.39	11.48	10.89	None
52	124.75	11.48	2.56	None

Max allowable Dc according to the defined chord length and radius of curvature is: 11.48 degree.

Curvatures are violated at (3) station points.

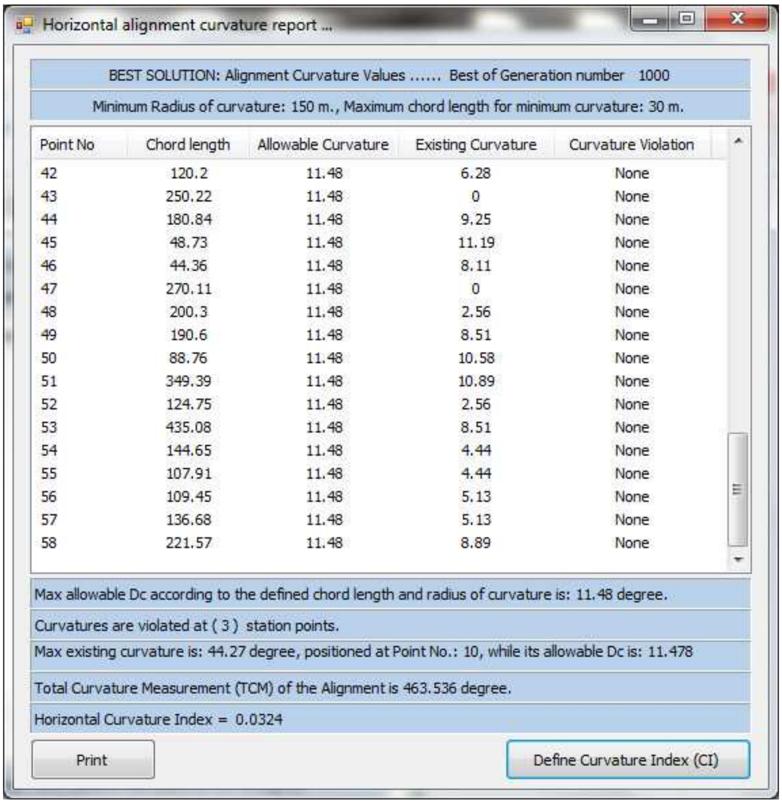
Max existing curvature is: 44.27 degree, positioned at Point No.: 10, while its allowable Dc is: 11.478

Total Curvature Measurement (TCM) of the Alignment is 463.536 degree.

Horizontal Curvature Index = 0.0324

Print | Define Curvature Index (CI)

Figure APP2.13: Horizontal curvature report output window (part 3)



Horizontal alignment curvature report ...

BEST SOLUTION: Alignment Curvature Values ...... Best of Generation number 1000

Minimum Radius of curvature: 150 m., Maximum chord length for minimum curvature: 30 m.

Point No	Chord length	Allowable Curvature	Existing Curvature	Curvature Violation
42	120.2	11.48	6.28	None
43	250.22	11.48	0	None
44	180.84	11.48	9.25	None
45	48.73	11.48	11.19	None
46	44.36	11.48	8.11	None
47	270.11	11.48	0	None
48	200.3	11.48	2.56	None
49	190.6	11.48	8.51	None
50	88.76	11.48	10.58	None
51	349.39	11.48	10.89	None
52	124.75	11.48	2.56	None
53	435.08	11.48	8.51	None
54	144.65	11.48	4.44	None
55	107.91	11.48	4.44	None
56	109.45	11.48	5.13	None
57	136.68	11.48	5.13	None
58	221.57	11.48	8.89	None

Max allowable Dc according to the defined chord length and radius of curvature is: 11.48 degree.

Curvatures are violated at (3) station points.

Max existing curvature is: 44.27 degree, positioned at Point No.: 10, while its allowable Dc is: 11.478

Total Curvature Measurement (TCM) of the Alignment is 463.536 degree.

Horizontal Curvature Index = 0.0324

Print | Define Curvature Index (CI)

Figure APP2.14: Horizontal curvature report output window (part 4)

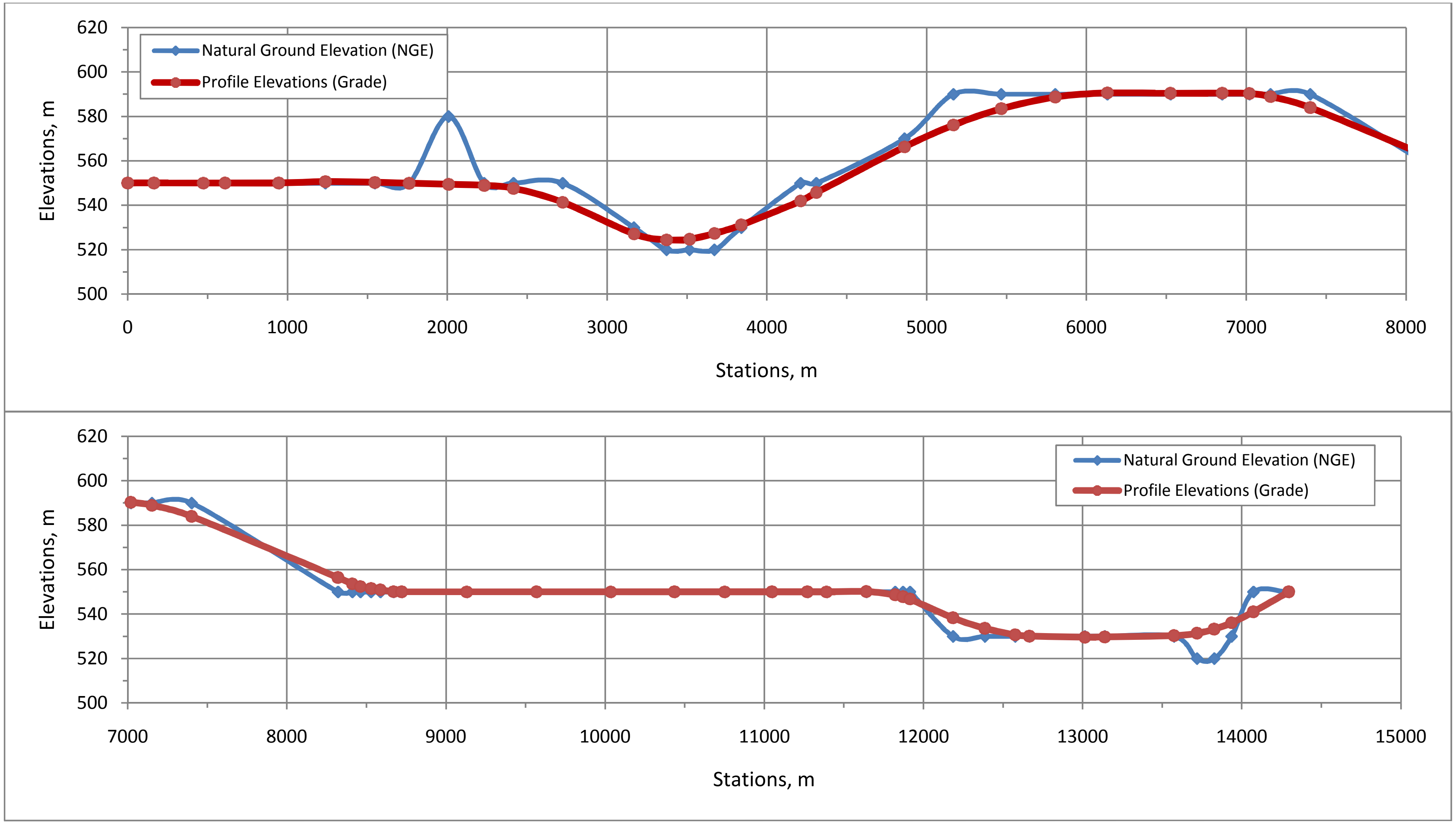
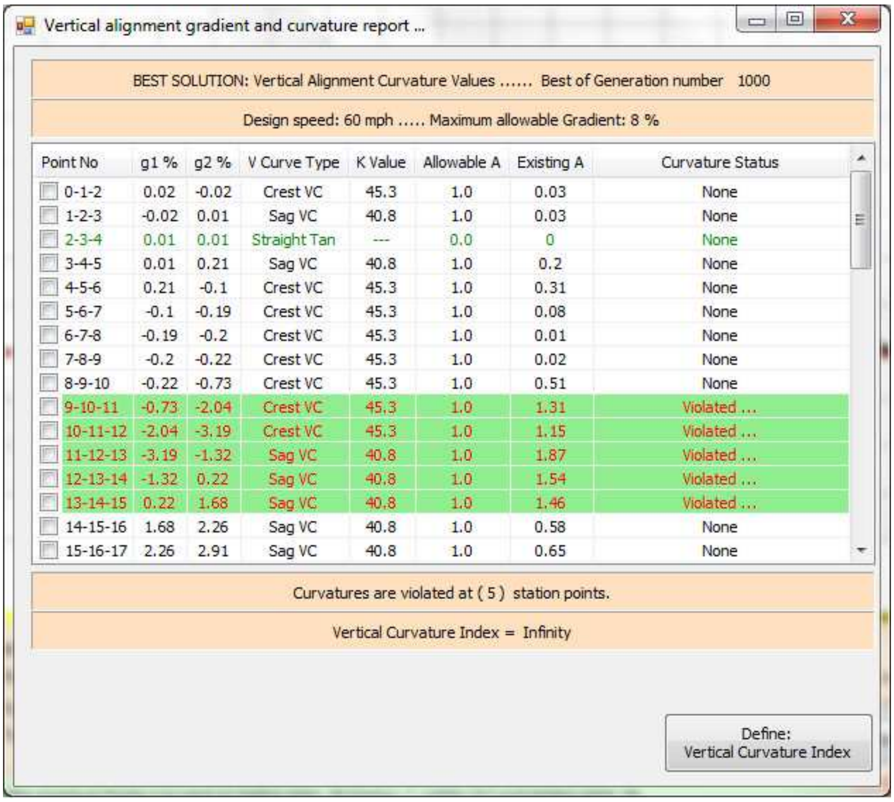


Figure APP2.15: The vertical alignment (alignment profile)



Vertical alignment gradient and curvature report ...

BEST SOLUTION: Vertical Alignment Curvature Values ...... Best of Generation number 1000

Design speed: 60 mph ..... Maximum allowable Gradient: 8 %

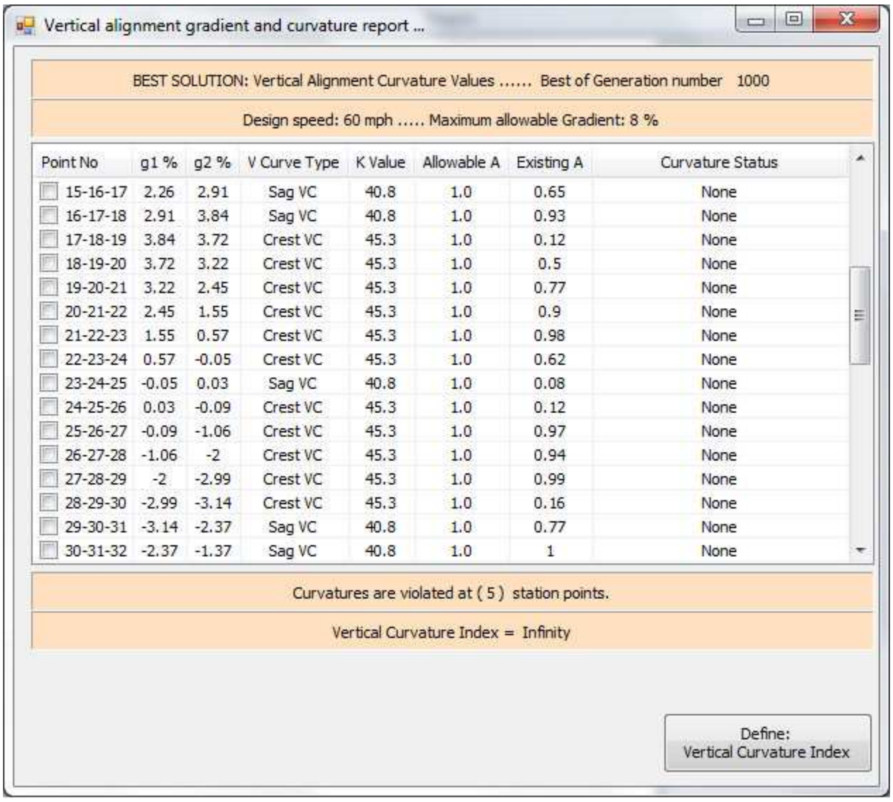
Point No	g1 %	g2 %	V Curve Type	K Value	Allowable A	Existing A	Curvature Status
0-1-2	0.02	-0.02	Crest VC	45.3	1.0	0.03	None
1-2-3	-0.02	0.01	Sag VC	40.8	1.0	0.03	None
2-3-4	0.01	0.01	Straight Tan	---	0.0	0	None
3-4-5	0.01	0.21	Sag VC	40.8	1.0	0.2	None
4-5-6	0.21	-0.1	Crest VC	45.3	1.0	0.31	None
5-6-7	-0.1	-0.19	Crest VC	45.3	1.0	0.08	None
6-7-8	-0.19	-0.2	Crest VC	45.3	1.0	0.01	None
7-8-9	-0.2	-0.22	Crest VC	45.3	1.0	0.02	None
8-9-10	-0.22	-0.73	Crest VC	45.3	1.0	0.51	None
9-10-11	-0.73	-2.04	Crest VC	45.3	1.0	1.31	Violated ...
10-11-12	-2.04	-3.19	Crest VC	45.3	1.0	1.15	Violated ...
11-12-13	-3.19	-1.32	Sag VC	40.8	1.0	1.87	Violated ...
12-13-14	-1.32	0.22	Sag VC	40.8	1.0	1.54	Violated ...
13-14-15	0.22	1.68	Sag VC	40.8	1.0	1.46	Violated ...
14-15-16	1.68	2.26	Sag VC	40.8	1.0	0.58	None
15-16-17	2.26	2.91	Sag VC	40.8	1.0	0.65	None

Curvatures are violated at (5) station points.

Vertical Curvature Index = Infinity

Define: Vertical Curvature Index

Figure APP2.16: Vertical curvature report output window (part 1)



Vertical alignment gradient and curvature report ...

BEST SOLUTION: Vertical Alignment Curvature Values ...... Best of Generation number 1000

Design speed: 60 mph ..... Maximum allowable Gradient: 8 %

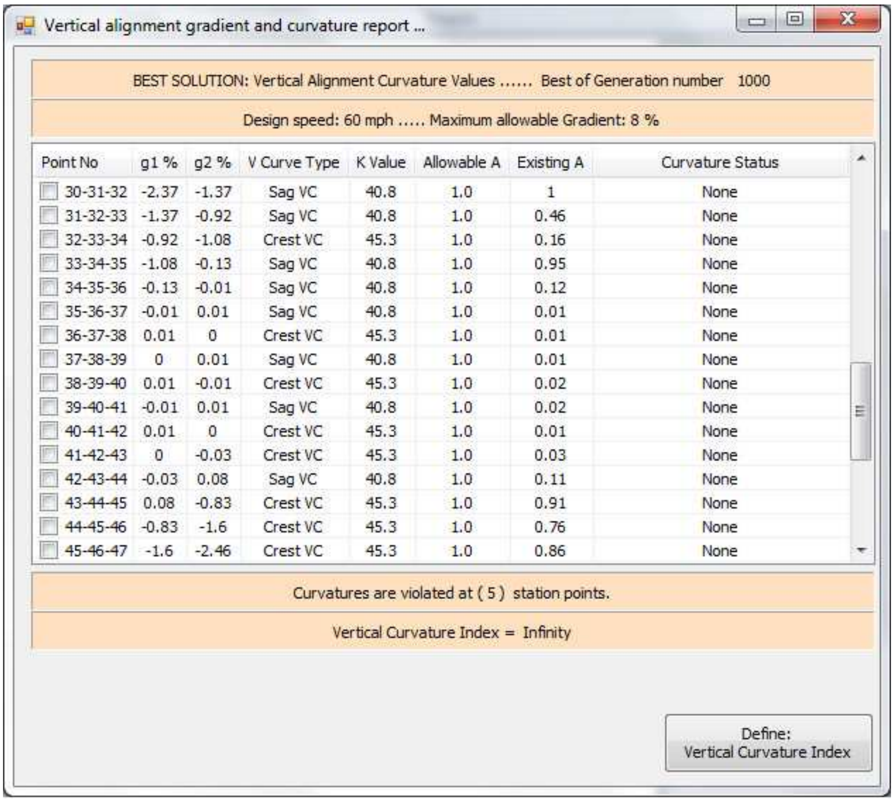
Point No	g1 %	g2 %	V Curve Type	K Value	Allowable A	Existing A	Curvature Status
15-16-17	2.26	2.91	Sag VC	40.8	1.0	0.65	None
16-17-18	2.91	3.84	Sag VC	40.8	1.0	0.93	None
17-18-19	3.84	3.72	Crest VC	45.3	1.0	0.12	None
18-19-20	3.72	3.22	Crest VC	45.3	1.0	0.5	None
19-20-21	3.22	2.45	Crest VC	45.3	1.0	0.77	None
20-21-22	2.45	1.55	Crest VC	45.3	1.0	0.9	None
21-22-23	1.55	0.57	Crest VC	45.3	1.0	0.98	None
22-23-24	0.57	-0.05	Crest VC	45.3	1.0	0.62	None
23-24-25	-0.05	0.03	Sag VC	40.8	1.0	0.08	None
24-25-26	0.03	-0.09	Crest VC	45.3	1.0	0.12	None
25-26-27	-0.09	-1.06	Crest VC	45.3	1.0	0.97	None
26-27-28	-1.06	-2	Crest VC	45.3	1.0	0.94	None
27-28-29	-2	-2.99	Crest VC	45.3	1.0	0.99	None
28-29-30	-2.99	-3.14	Crest VC	45.3	1.0	0.16	None
29-30-31	-3.14	-2.37	Sag VC	40.8	1.0	0.77	None
30-31-32	-2.37	-1.37	Sag VC	40.8	1.0	1	None

Curvatures are violated at (5) station points.

Vertical Curvature Index = Infinity

Define: Vertical Curvature Index

Figure APP2.17: Vertical curvature report output window (part 2)



Vertical alignment gradient and curvature report ...

BEST SOLUTION: Vertical Alignment Curvature Values ...... Best of Generation number 1000

Design speed: 60 mph ..... Maximum allowable Gradient: 8 %

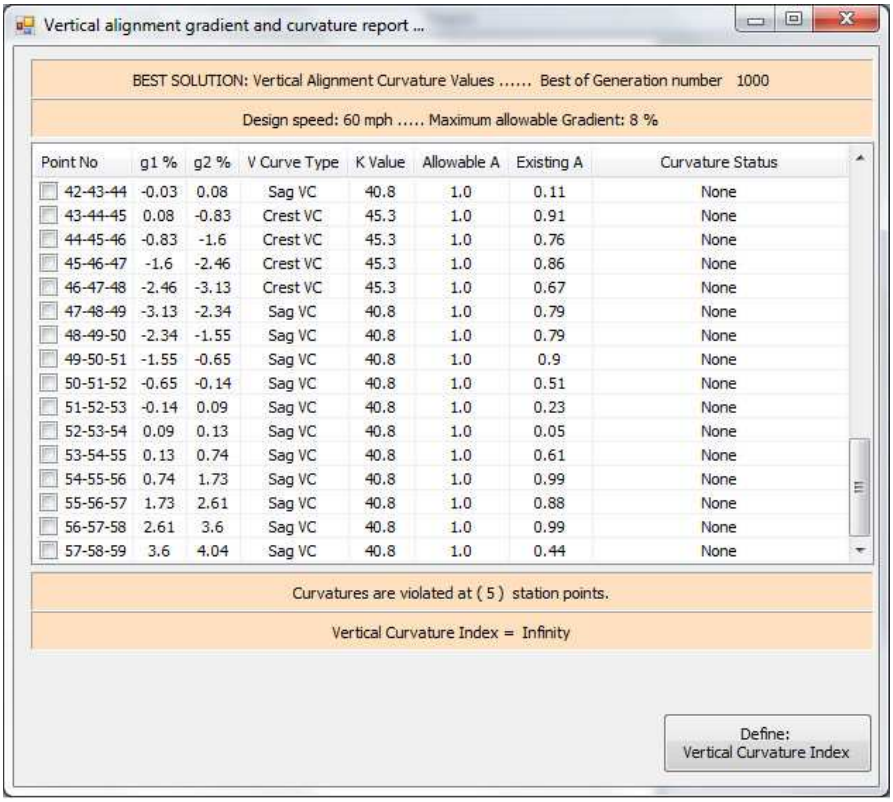
Point No	g1 %	g2 %	V Curve Type	K Value	Allowable A	Existing A	Curvature Status
30-31-32	-2.37	-1.37	Sag VC	40.8	1.0	1	None
31-32-33	-1.37	-0.92	Sag VC	40.8	1.0	0.46	None
32-33-34	-0.92	-1.08	Crest VC	45.3	1.0	0.16	None
33-34-35	-1.08	-0.13	Sag VC	40.8	1.0	0.95	None
34-35-36	-0.13	-0.01	Sag VC	40.8	1.0	0.12	None
35-36-37	-0.01	0.01	Sag VC	40.8	1.0	0.01	None
36-37-38	0.01	0	Crest VC	45.3	1.0	0.01	None
37-38-39	0	0.01	Sag VC	40.8	1.0	0.01	None
38-39-40	0.01	-0.01	Crest VC	45.3	1.0	0.02	None
39-40-41	-0.01	0.01	Sag VC	40.8	1.0	0.02	None
40-41-42	0.01	0	Crest VC	45.3	1.0	0.01	None
41-42-43	0	-0.03	Crest VC	45.3	1.0	0.03	None
42-43-44	-0.03	0.08	Sag VC	40.8	1.0	0.11	None
43-44-45	0.08	-0.83	Crest VC	45.3	1.0	0.91	None
44-45-46	-0.83	-1.6	Crest VC	45.3	1.0	0.76	None
45-46-47	-1.6	-2.46	Crest VC	45.3	1.0	0.86	None

Curvatures are violated at (5) station points.

Vertical Curvature Index = Infinity

Define: Vertical Curvature Index

Figure APP2.18: Vertical curvature report output window (part 3)



Vertical alignment gradient and curvature report ...

BEST SOLUTION: Vertical Alignment Curvature Values ...... Best of Generation number 1000

Design speed: 60 mph ..... Maximum allowable Gradient: 8 %

Point No	g1 %	g2 %	V Curve Type	K Value	Allowable A	Existing A	Curvature Status
42-43-44	-0.03	0.08	Sag VC	40.8	1.0	0.11	None
43-44-45	0.08	-0.83	Crest VC	45.3	1.0	0.91	None
44-45-46	-0.83	-1.6	Crest VC	45.3	1.0	0.76	None
45-46-47	-1.6	-2.46	Crest VC	45.3	1.0	0.86	None
46-47-48	-2.46	-3.13	Crest VC	45.3	1.0	0.67	None
47-48-49	-3.13	-2.34	Sag VC	40.8	1.0	0.79	None
48-49-50	-2.34	-1.55	Sag VC	40.8	1.0	0.79	None
49-50-51	-1.55	-0.65	Sag VC	40.8	1.0	0.9	None
50-51-52	-0.65	-0.14	Sag VC	40.8	1.0	0.51	None
51-52-53	-0.14	0.09	Sag VC	40.8	1.0	0.23	None
52-53-54	0.09	0.13	Sag VC	40.8	1.0	0.05	None
53-54-55	0.13	0.74	Sag VC	40.8	1.0	0.61	None
54-55-56	0.74	1.73	Sag VC	40.8	1.0	0.99	None
55-56-57	1.73	2.61	Sag VC	40.8	1.0	0.88	None
56-57-58	2.61	3.6	Sag VC	40.8	1.0	0.99	None
57-58-59	3.6	4.04	Sag VC	40.8	1.0	0.44	None

Curvatures are violated at (5) station points.

Vertical Curvature Index = Infinity

Define: Vertical Curvature Index

Figure APP2.19: Vertical curvature report output window (part 4)

APPENDIX 3

A SPECIFIC WORLD CASE FOR A 3D HIGHWAY ALIGNMENT TEST

A test scenario was conducted for a 3D highway alignment optimisation on a world similar to the World W_5 used for 2D highway alignment; but, instead of using different land features to format the inversed F shaped lands (the zigzagged areas) as land use features, the same form was reproduced using high elevation lands. In other words, the world was specifically designed as a 3D format with unique land features (equal land unit costs) and high elevations in forms of two inversed F shapes.

The test was conducted using the parameters of scenario S7 from Table 5.5. No constraints were considered and the results showed the capability of the model to respond for the same alignment configuration as obtained when the model tested with F shaped land categories. The world features and the results were as shown in Figures APP3.1 to APP3.5

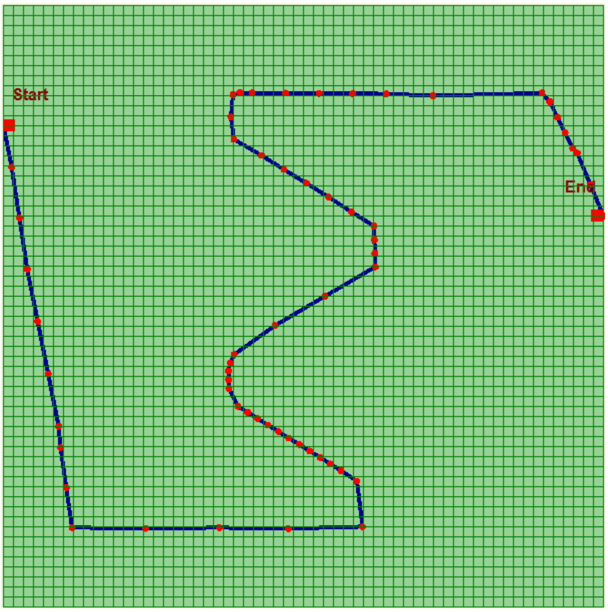


Figure APP3.1: The horizontal alignment on a land use map

(Note that the lands have unique characteristics with the same unit cost)

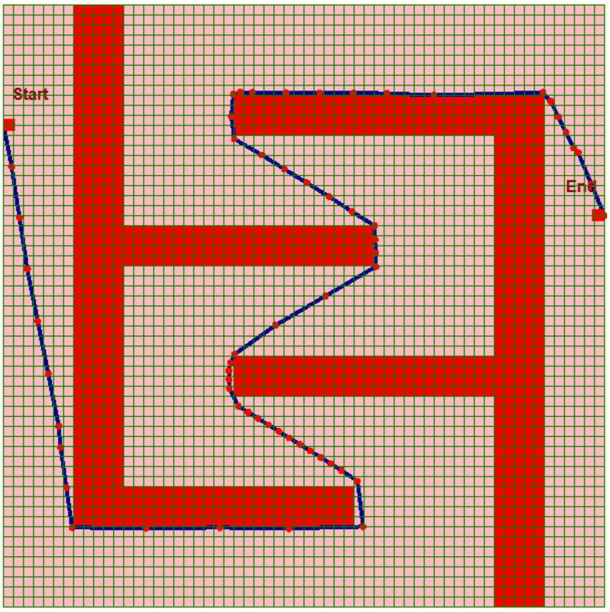


Figure APP3.2: The horizontal alignment on a digital elevation map

(The darker the cell the higher is the elevation)

(Note: Two elevation values used to format the topography of this world. The light red colour cells are with elevations 555m and the dark red cells are with elevations 888m)

Figure APP3.3 shows the topography of the area (surface terrain).

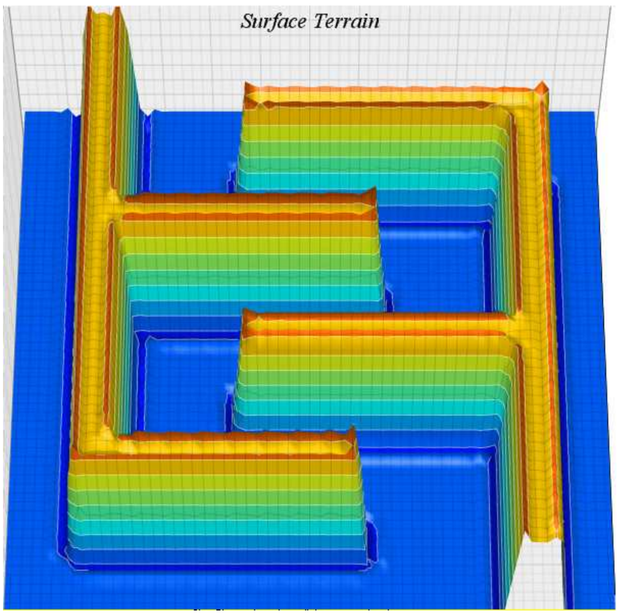


Figure APP3.3: The topography of an area with two inversed F shaped high elevations

(surface terrain)

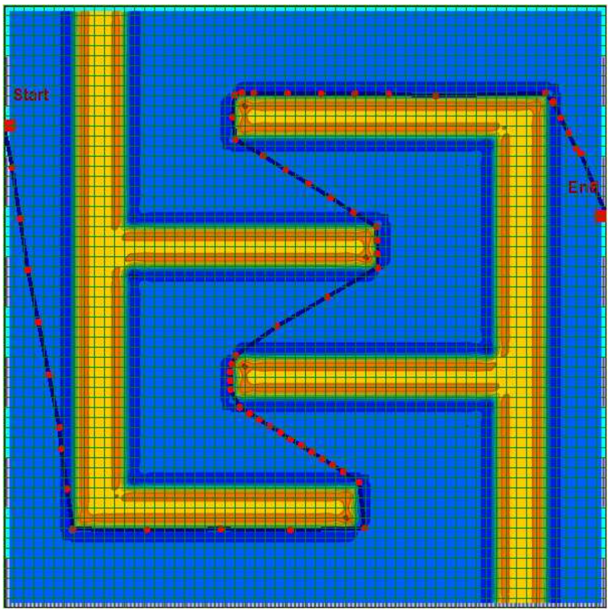


Figure APP3.4: The horizontal alignment on a contour map

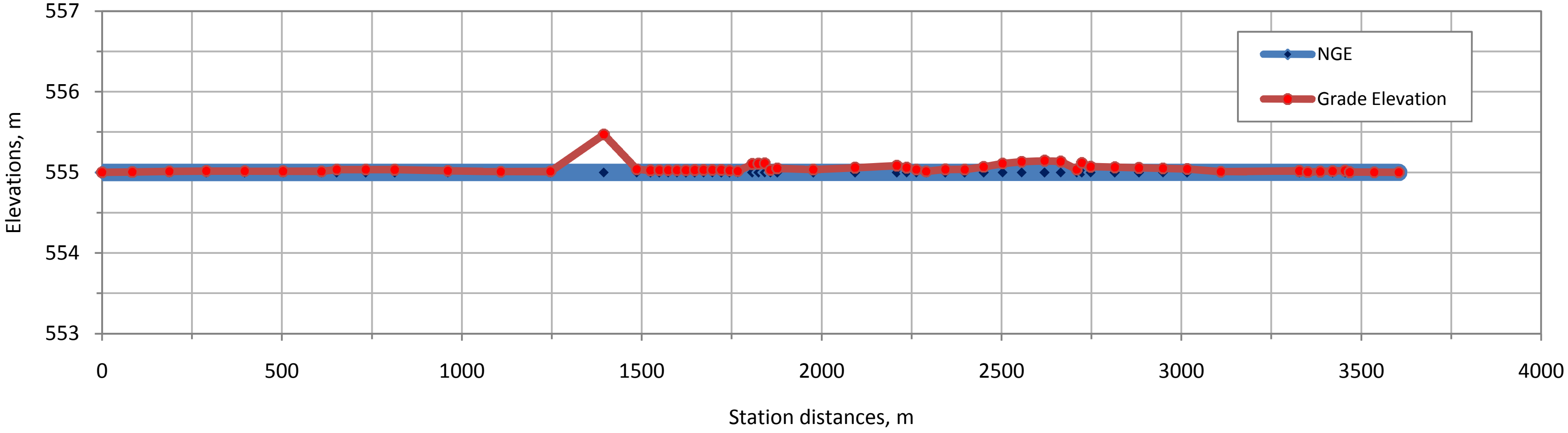


Figure APP3. 5: The vertical alignment (profile)

Table APP3.1 lists the station distances and the elevations of the vertical alignment solution along the centreline of the alignment.

Table APP3.1: The elevations of the alignment profile along the alignment centreline

Stations	NGE	Grade Elevation	Stations	NGE	Grade Elevation
0.00	555	555.00	1,876.80	555	555.05
84.23	555	555.01	1,976.95	555	555.03
186.94	555	555.01	2,093.04	555	555.06
289.64	555	555.02	2,209.13	555	555.08
396.40	555	555.02	2,236.32	555	555.06
503.15	555	555.01	2,263.51	555	555.04
609.90	555	555.01	2,290.69	555	555.01
652.12	555	555.03	2,343.79	555	555.04
732.84	555	555.03	2,397.62	555	555.03
813.55	555	555.03	2,450.47	555	555.07
961.14	555	555.02	2,503.07	555	555.11
1,108.35	555	555.01	2,556.16	555	555.13
1,246.16	555	555.01	2,619.67	555	555.14
1,394.59	555	555.47	2,664.56	555	555.13
1,486.05	555	555.04	2,709.14	555	555.04
1,524.34	555	555.02	2,723.31	555	555.12
1,548.93	555	555.03	2,747.88	555	555.07
1,573.52	555	555.03	2,814.93	555	555.06
1,598.12	555	555.03	2,881.98	555	555.06
1,622.71	555	555.03	2,949.03	555	555.05
1,647.30	555	555.03	3,016.08	555	555.04
1,671.90	555	555.03	3,109.72	555	555.01
1,696.49	555	555.03	3,327.86	555	555.02
1,721.08	555	555.03	3,351.30	555	555.01
1,744.10	555	555.02	3,385.82	555	555.01
1,767.11	555	555.02	3,420.34	555	555.02
1,807.22	555	555.10	3,454.87	555	555.02
1,824.51	555	555.11	3,468.13	555	555.00
1,841.81	555	555.11	3,536.12	555	555.00
1,858.04	555	555.03	3,604.11	555	555.00

APPENDIX 4

TESTS FOR DIFFERENT STATION POINT NUMBERS

It was shown that the number of the station points were problem dependent. Irregular areas and the requirement for a more precise alignment representation demanded larger station point numbers.

A test scenario was prepared using the same parameters as used in Table 6.3 (S1) to investigate the implications of small numbers of station points for relatively irregular areas. For this purpose the same world W_9 was used but instead of using 60 station points; 20 were used. The optimum alignment results was a solution with fitness 13,084,165 (which was optimum as 42.3% of the most optimum solution with fitness 8,294,256), passed through high cost areas, and violated horizontal and vertical curvatures at many locations (see Figures APP4.1 to APP4.5).

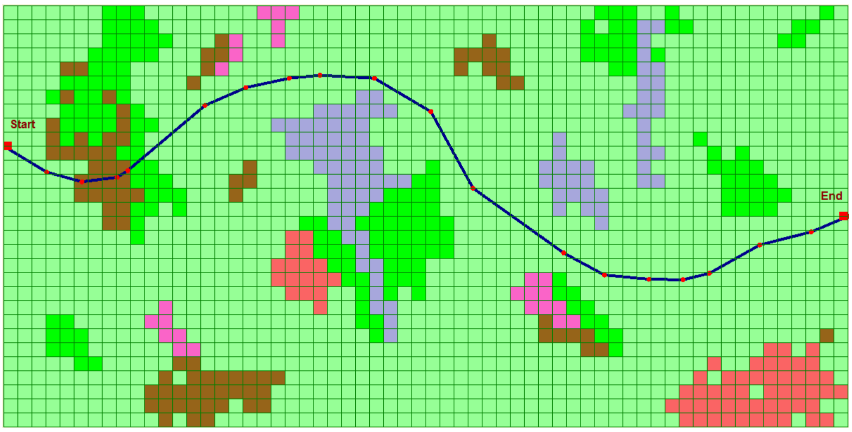


Figure APP4.1: An optimum solution configured using 20 station points (land use map)

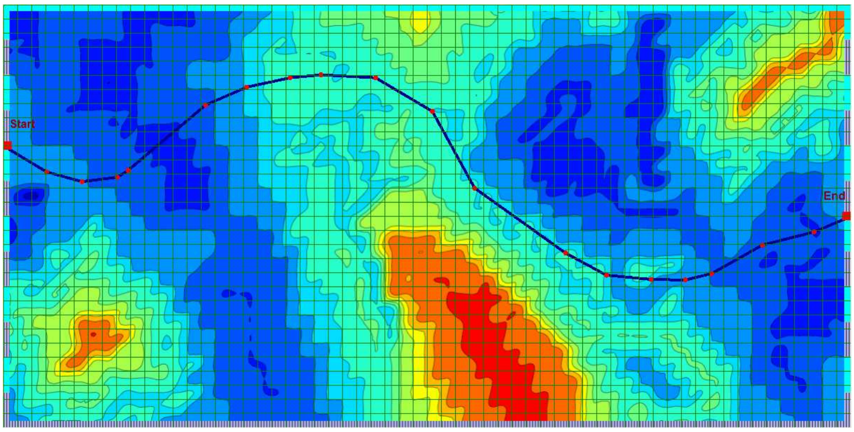
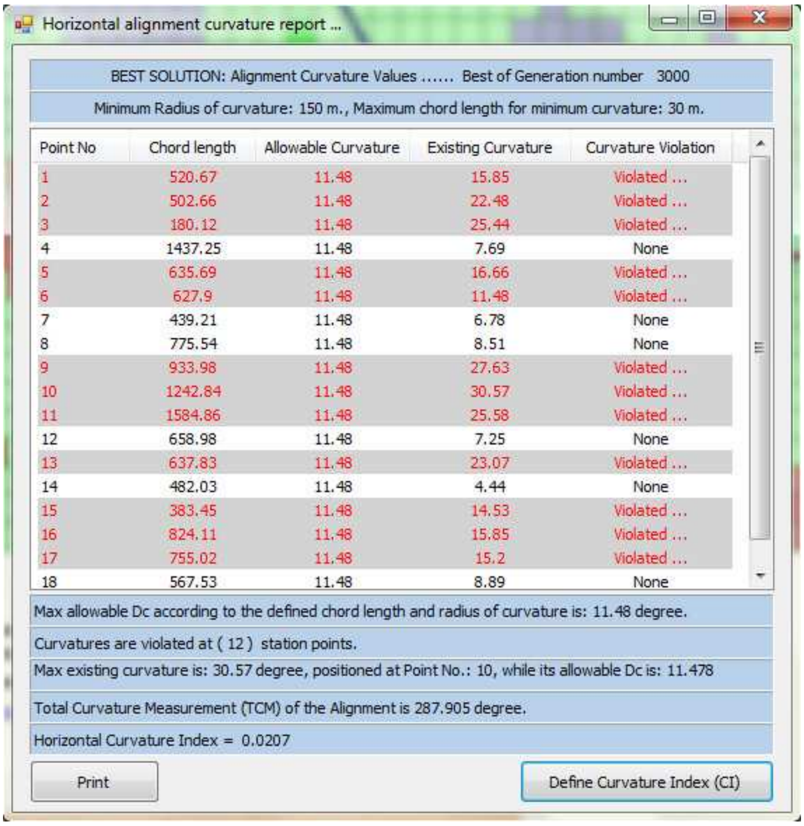


Figure APP4.2: An optimum solution configured using 20 station points (contour map)


Horizontal alignment curvature report ...

BEST SOLUTION: Alignment Curvature Values ...... Best of Generation number 3000

Minimum Radius of curvature: 150 m., Maximum chord length for minimum curvature: 30 m.

Point No	Chord length	Allowable Curvature	Existing Curvature	Curvature Violation
1	520.67	11.48	15.85	Violated ...
2	502.66	11.48	22.48	Violated ...
3	180.12	11.48	25.44	Violated ...
4	1437.25	11.48	7.69	None
5	635.69	11.48	16.66	Violated ...
6	627.9	11.48	11.48	Violated ...
7	439.21	11.48	6.78	None
8	775.54	11.48	8.51	None
9	933.98	11.48	27.63	Violated ...
10	1242.84	11.48	30.57	Violated ...
11	1584.86	11.48	25.58	Violated ...
12	658.98	11.48	7.25	None
13	637.83	11.48	23.07	Violated ...
14	482.03	11.48	4.44	None
15	383.45	11.48	14.53	Violated ...
16	824.11	11.48	15.85	Violated ...
17	755.02	11.48	15.2	Violated ...
18	567.53	11.48	8.89	None

Max allowable Dc according to the defined chord length and radius of curvature is: 11.48 degree.

Curvatures are violated at (12) station points.

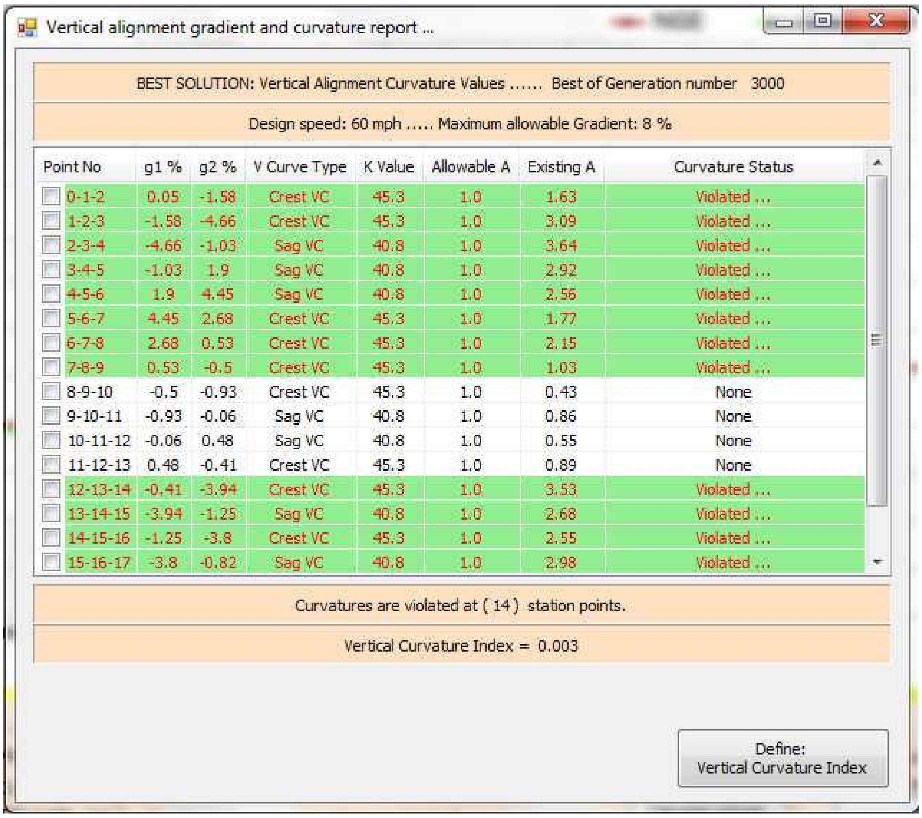
Max existing curvature is: 30.57 degree, positioned at Point No.: 10, while its allowable Dc is: 11.478

Total Curvature Measurement (TCM) of the Alignment is 287.905 degree.

Horizontal Curvature Index = 0.0207

Print | Define Curvature Index (CI)

Figure APP4.3: Horizontal curvature status report at the station point locations



Point No	g1 %	g2 %	V Curve Type	K Value	Allowable A	Existing A	Curvature Status
0-1-2	0.05	-1.58	Crest VC	45.3	1.0	1.63	Violated ...
1-2-3	-1.58	-4.66	Crest VC	45.3	1.0	3.09	Violated ...
2-3-4	-4.66	-1.03	Sag VC	40.8	1.0	3.64	Violated ...
3-4-5	-1.03	1.9	Sag VC	40.8	1.0	2.92	Violated ...
4-5-6	1.9	4.45	Sag VC	40.8	1.0	2.56	Violated ...
5-6-7	4.45	2.68	Crest VC	45.3	1.0	1.77	Violated ...
6-7-8	2.68	0.53	Crest VC	45.3	1.0	2.15	Violated ...
7-8-9	0.53	-0.5	Crest VC	45.3	1.0	1.03	Violated ...
8-9-10	-0.5	-0.93	Crest VC	45.3	1.0	0.43	None
9-10-11	-0.93	-0.06	Sag VC	40.8	1.0	0.86	None
10-11-12	-0.06	0.48	Sag VC	40.8	1.0	0.55	None
11-12-13	0.48	-0.41	Crest VC	45.3	1.0	0.89	None
12-13-14	-0.41	-3.94	Crest VC	45.3	1.0	3.53	Violated ...
13-14-15	-3.94	-1.25	Sag VC	40.8	1.0	2.68	Violated ...
14-15-16	-1.25	-3.8	Crest VC	45.3	1.0	2.55	Violated ...
15-16-17	-3.8	-0.82	Sag VC	40.8	1.0	2.98	Violated ...

Figure APP4.4: Vertical curvature status report at the station point locations

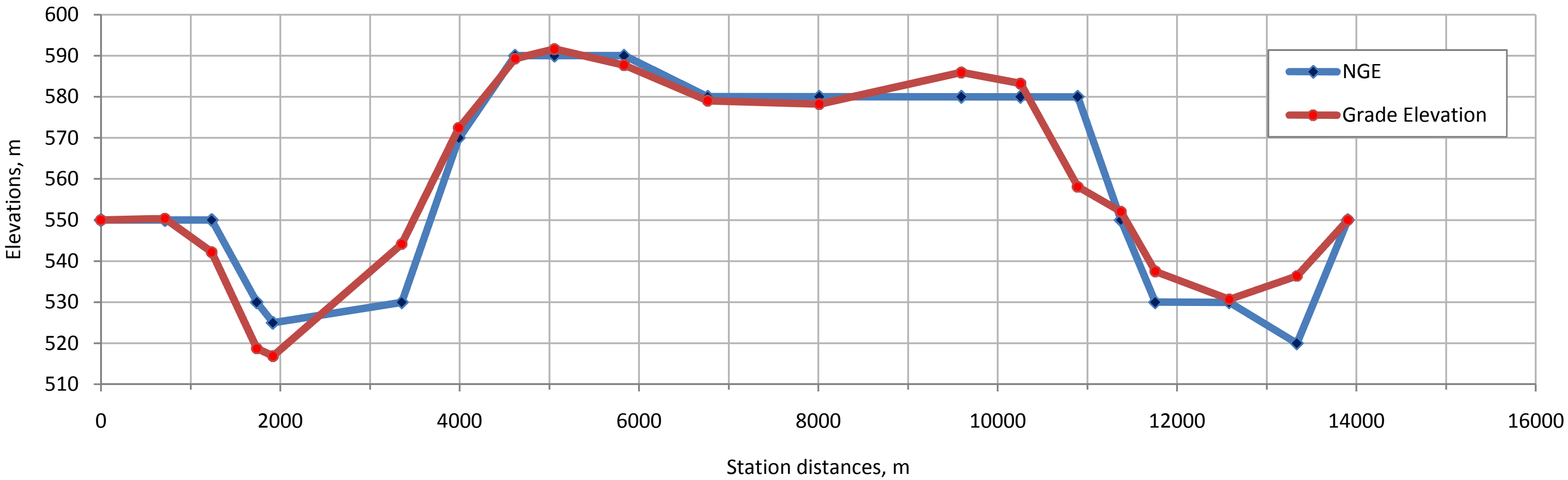


Figure APP4.5: The vertical Alignment (Profile)

APPENDIX 5

VERIFICATION TESTS FOR DIFFERENT POPULATION SIZES

Two additional test scenarios were made to show how the model behaves with different population sizes. This could be regarded as a proof for what have been mentioned (proved) in the thesis body.

The tests were reproduced using 500 and 7500 population sizes. The results were compared with those found when 5000 population size was used. The same scenario as S1 (see Table 6.3 Chapter 6) was used and the results were compared.

Figures APP5.1 and APP5.2 shows the results for 500 population size. The solution fitness was 11,137,494 (65.7 % of the most optimum solution with fitness 8,294,256).

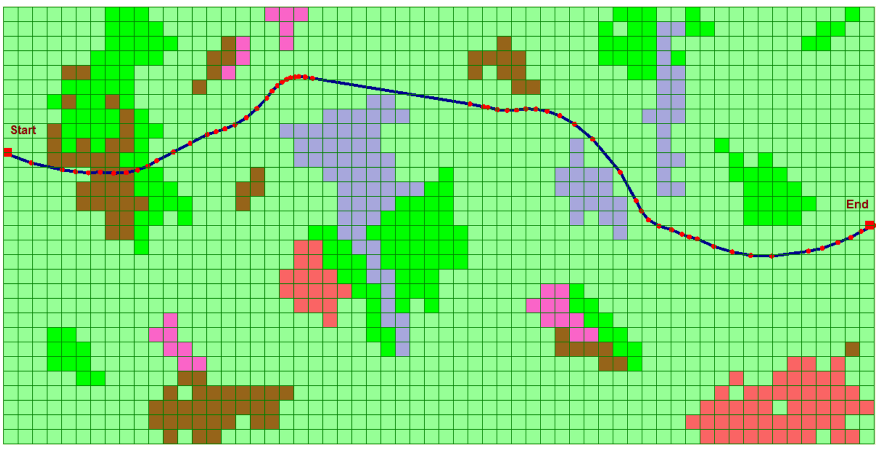


Figure APP5.1: An optimum solution when 500 individuals are used (Land use map)

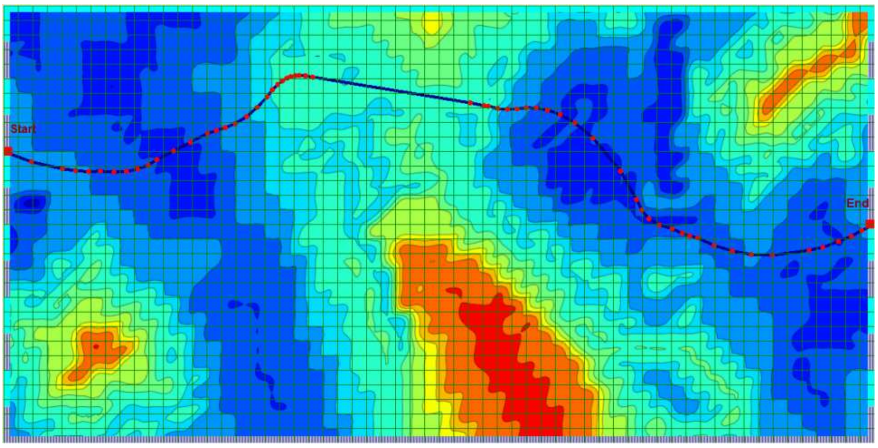


Figure APP5.2: An optimum solution when 500 individuals are used (Contour map)

Figures APP5.3 and APP5.4 shows the results for 7500 population size. The solution fitness was 8,522,688 (97.5 % of the most optimum solution found with fitness 8,294,256).

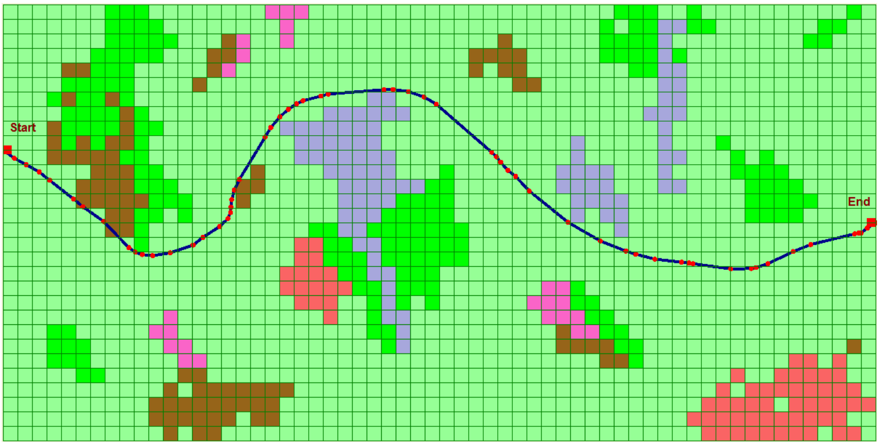


Figure APP5.3: An optimum solution using 7500 individuals as the population size (Land use map)

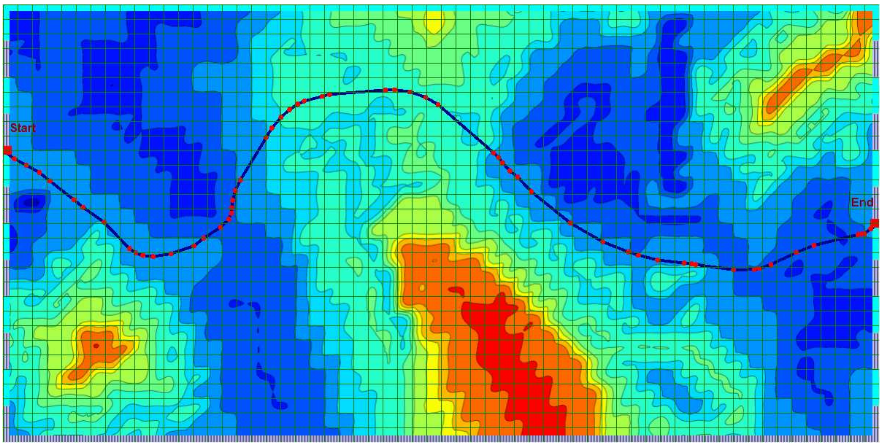


Figure APP5.4: An optimum solution using 7500 individuals as the population size

(Contour map)

It can be seen that population sizes less than 5000 resulted in producing a local optimum solution while more than 5000 resulted in nearly the same solution as with 5000.

It should be noted that these tests were specific to the sizes and the irregularity of the worlds used.