GASNV Environment: Building Virtual Sewer Networks Optimized by Genetic Algorithms

Andre Bevilaqua[1], Laurence Rodrigues do Amaral[2], Marcos Wagner de Souza Ribeiro[1]
[1]Department of Computer Sciences
Federal University of Goias/Jatai
Jatai, Brazil
[2]Faculty of Computation
Federal University of Uberlandia
Patos de Minas, Brazil
{lramaral, marcos_wagner}@yahoo.com.br, andrebevilaquax@gmail.com

ABSTRACT: *Less than two thirds of the world population have access to improved sanitation facilities. This number shows a great disparity between regions when looking at the global picture. The most alarming problem is observed in Southern Asia, followed by Eastern Asia and Sub-Saharan Africa. There is a wide variety of sanitation problems being faced in the world and one of the most important among them is the lack of piped sewer systems. A lot of attention was focused on the optimal design of storm sewer networks in the past decades. Storm water networks, specifically, can be considered an essential part of the infrastructure of any society. Every aspect involved in the construction and maintenance of these networks requires a huge amount of investment. In general, the lack of sanitation facilities occurs due to the high cost involved at the implantation. The present work proposes a computational evolutionary environment aiming to provide an optimal decision regarding the implantation of piped sewer systems. The method uses Genetic Algorithms and Information Visualization concepts, and has as a main goal presenting an alternative method that may be used to improve the sanitation statistics by covering a larger area with piped sewer systems, and most of all, reducing the costs and impact of the implantation.*

Keywords: Genetic Algorithm, Information Visualization, Sanitation, Piped Sewer Systems

Received: 23 July 2012, Revised 20 September 2012, Accepted 1 October 2012

1. Introduction

Less than two thirds of the world population has access to improved sanitation facilities. Among the 2.6 billion people in the world who do not use improved sanitation facilities, by far the greatest number are in Southern Asia, but there are also large numbers in Eastern Asia and Sub-Saharan Africa [4].

The Figure 1 gives a better idea of the world regions with the most number of people without access to improved sanitation facilities. Based on provided data, China and India together are home to more than a third of all the world population. Even with so many habitants, these two countries still need great improvement on sanitation facilities. Statistics show that the proportion of the population using improved sanitation facilities in China is 55%, while in India the value is 31% - and they are just two among a lot of other examples of countries which face difficulties on the referred area, making this a problem that deserves priority attention in the world. The number of families without access to adequate waste disposal in poor urban and rural areas continues to grow in manyplaces, and to change this situation a lot of study regarding new strategies and tools to be applied is needed [4].

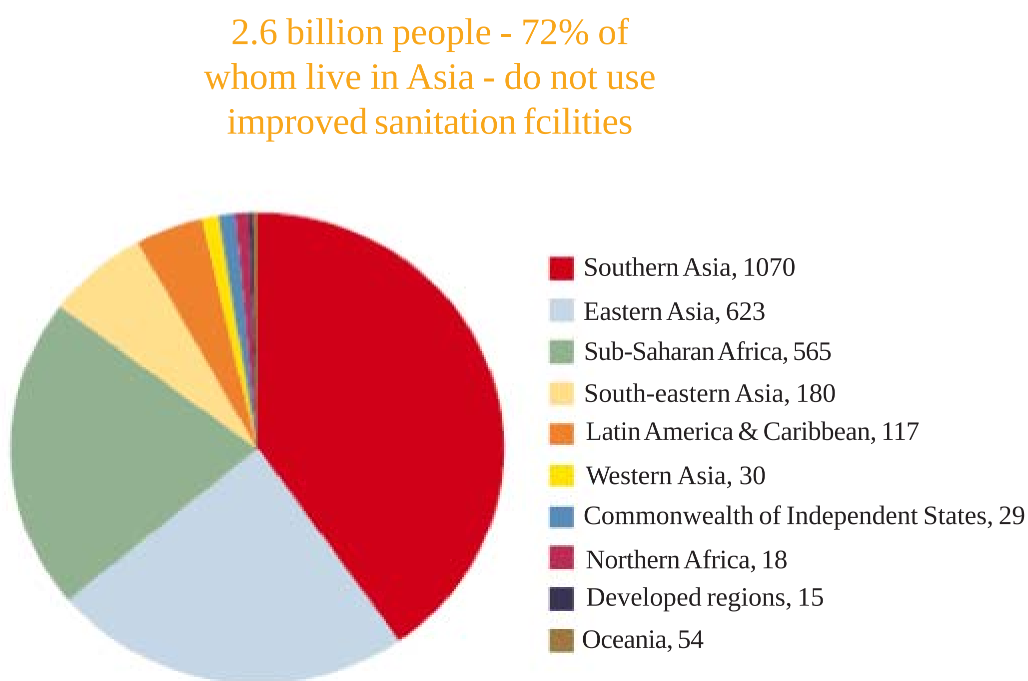

Figure 1. Regional Distribution of the 2.6 billions of people without access to improved sanitation facilities in 2008, population (million)[4]

2. Strategies and Sanitation Infrastructure

Storm water networks, specifically, can be considered an essential part of the infrastructure of any society. Every aspect involved in the construction and maintenance of these networks always requires a huge amount of investment. The great scope of the problem makes any reduction of total cost in the construction - through proper design - result in considerable savings. Pipes and excavations constitute the main cost of storm water networks, but the relation between them is a bit contradictious. When some strategies are applied aiming to reduce the pipe size, for example, there is a chance that the excavation costs will need to be higher, so any economical design requires an optimal trade-o between the excavation and pipes costs, which cannot be achieved by engineering judgments [1].

A lot of attention has been focused on the optimal design of storm sewer networks in the past decades. According to [1], the attempts for optimization of storm water networks can be classified, based on the kind of programming techniques used, in three groups.

2.1 Dynamic Programming (DP)

The methods that belongs to this group are the most frequently used to optimize the implantation of sewer networks. They are known for being capable of finding global optimal solution, but are not applicable to real-world sewer networks because of their inability to deal with the great dimensionality involved on this problem .

2.2 Linear Programming (LP)

Some approaches using linear programming were also tested. Normally the results of LP methods were not so effective, and in order to have better results, researchers tried to use some heuristics along with them, being this strategy responsible for good results .

2.3 Evolutionary Computation (EC)

Evolutionary Computation methods, particularly Genetic Algorithms (GA), have been used in many areas of the water resource industry . According to [1], the GA method obtained good results in works involving several areas of the water resource industry. In this work, the most important fact to be emphasized is the success obtained by the method when used to optimize piped sewer networks.

After all the information (about the difficulties faced in sanitation area all over the world) given until now we mayinfer that a technique related to optimization of piped sewer networks could be well accepted by researchers and companies.

This paper has as its main goal to show a system developed to assist the optimized planning of piped sewer networks using Genetic Algorithms (GA's) and Information Visualization (IV). Some important IV characteristics considered very useful to the purpose of this paper are shown at the following section.

3. Information Visualization (IV)

Every Information Visualization method, from the less to the more complex, has a common goal: help you see what the data has to say [5].

When someone receives a highly complex information, whatever is the reason of the complexity, the first and moreclassical reaction is to lose interest on it [5]. Data can't be considered just a bucket of information; the way you present this information depends on a context that must be well studied so as to provide the user the best understanding.

Graphics, for example, are an important Information Visualization method. All the colors, position of the information,captions and points showed on a graphic must be created in order to give the user a more clear and concise analysis of the data [5].

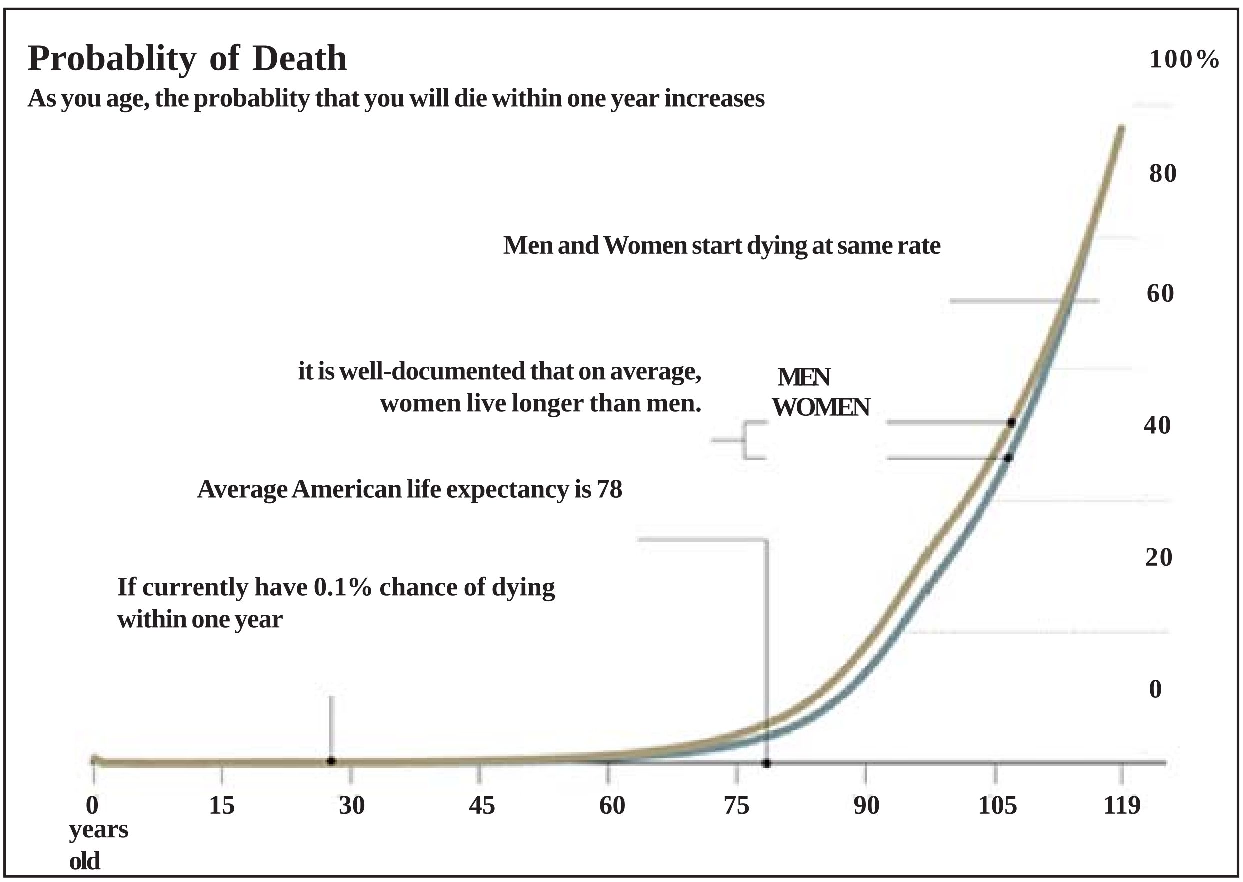

Figure 2. Graphic showing the probability of death depending on a person age[5]

The Figure 2 shows an example of how a graphic can be used in order to present an information in a synthesized and clear way - in this case, the probability of a person's death based on his/her age. This type of visualization was considered by us a very good alternative to show the estimates proposed by the system presented on this paper.

Simulation visualization seeks to convert data and behavior of a system, like the GASNV (Genetic Algorithm Sewer Network Visualization) presented on this paper, making the information more understandable and user-friendly. Using this "*improved* " information, users have the possibility to analyze the system behavior, the parameters and dimensioning, and have a better understanding of the possibility to apply it on a real situation [3].

Information Visualization using virtual reality broke the limit of screen size and created the possibility of viewing data in 3-D environments generated by computer. Besides the advantage of spatial view, this new visualization paradigm allows objects to be represented as they really are. In multimedia environments, the objects were represented by symbols, such as buttons, menus, icons, etc [3].

Computer simulation can be defined as a process of designing a model. In this paper, two models (2D and 3D) were created, aiming to provide the best visualization for the user of the real or theoretical system executing it in a computer and analyzing the output data[3].

After showing this overview about the advantages provided by IV methods, it is important to stress that these concepts were considered very useful to improve users understanding of the data generated by the Genetic Algorithm, being this the main reason that motivated the design of the 2D and 3D models used on GASNV system.

4. Evoulutionary Environment

4.1 Introduction

Genetic Algorithms (GA's) can be defined as computational search methods based on natural evolution and genetic mechanisms, simulating Darwin's natural selection theory. They can be defined as Evolutionary Computation approaches and can be considered a subarea of Artificial Intelligence focused on the study of computational methods based on Darwin's evolutionary theory. GA's are also parallel algorithms which manipulate a group (population) of individuals that represents chromosomes and by applying reproduction, survival, and other operators, transforms the current population in a new one, always aiming at creating a group of individuals better than before, based on a fitness function [2].

Genetic Algorithms are normally used in constrained optimization problems. When facing unconstrained problems(such as sewer networks) it requires a transformation of the underlying constrained problem to an unconstrained one. A strategy which is usually used to perform this transformation are the penalty methods, in which the constraints are included in the objective function via a penalty cost term, resulting in the following penalized form of the objective function [1].

4.2 Individual Representation

The first step when creating a GA model is to define the chromosome structure that will be used to represent the problem.

The system proposed in this paper, named GASNV (Genetic Algorithm Sewer Network Visualization), has the following chromosome codification:

The individual used in the GASNV was defined as a group of points (cartesian coordinates). These coordinates are taken from a 2D map that is modeled based on a real world city (the system examples showed in this paper used as model just a neighborhood of a real city) where the optimization is needed. The GA always takes the information needed to execute from the 2D model. Initial and final points are defined by the user on the model and the GA execution returns the shorter path between these two points allied with the smaller impact factor (paved, unpaved, historical preservation and public utility). This path is where the main pipe (the one with the greater flow) will be installed.

4.2.1 Initial Population

The initial population is a group of individuals (points of the 2D model). These individuals are generated randomly, but they must represent a valid path between the initial and final points defined by the user. The algorithm stops the initial population generation when all the individuals needed are created and valid. The Figure 3 represents the chromosome structure. Each "*P*" value corresponds to a point of the map and for each one of them there is another vector of possibleconnections and an impact factor (type of restriction) of the corresponding connection.

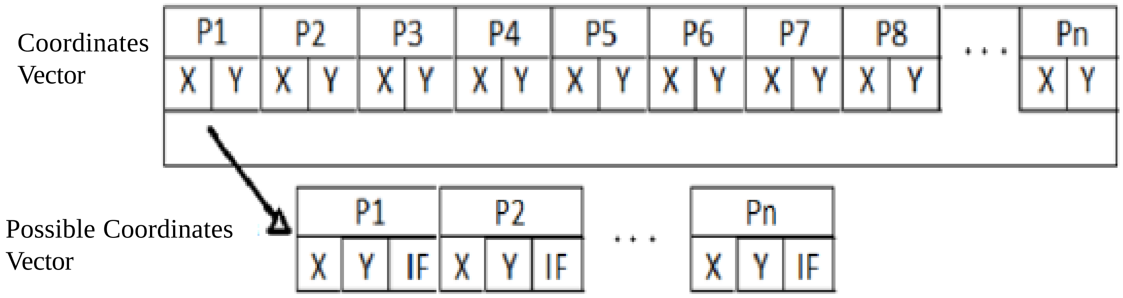

Figure 3. Chromosome Codification

The individual generated by the GASNV has a group of coordinates that represents the main pipe of the network, like shown on the Figure 4. When an individual is chosen as the best individual, all the points between the initial and final points defined by the user that don't belong to the individual are connected to the main pipe, by the genetic algorithm, creating the sub-mesh auxiliary network) of the project. The Figure 5 shows a didactic example of a main pipe path proposed by the GASNV taking the initial and final points chosen by the user.

P1		P2		P3		P4		P5		P6		P7		P8	
0	0	1	0	1	1	1	2	1	3	2	3	3	3	3	4

Figure 4. Individual Example

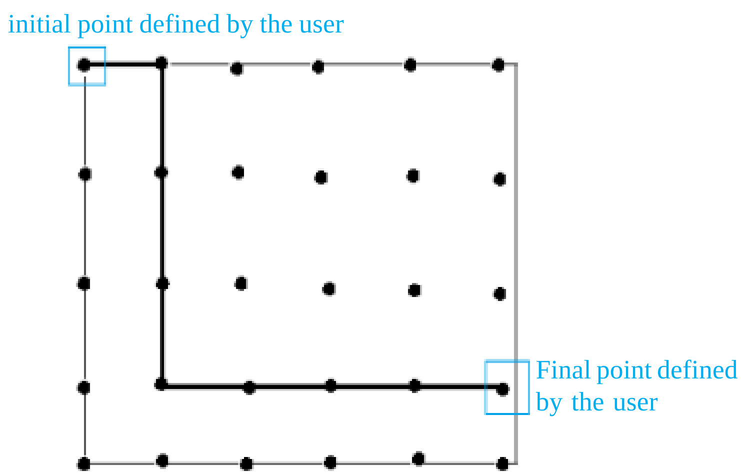

Figure 5. Individual outline

4.2.2 Fitness Function

The fitness function used on the GASNV is shown in the Equation 1. During the execution, for each individual, the system calculates the distance between the connection points and multiplies for the stretch impact factor. GASNV evaluates all the individuals in each generation and the one with the bigger fitness value is chosen as the pathwhere the main pipe will be installed.

$$Fitness\ Function = \frac{1}{Distance * IF} \qquad (1)$$

An individual fitness is calculated based on two variables, "*distance*" and "*impact factor*". To obtain the distance between each connected points of an individual the function 2 was used. It is important to remember that this function was necessary because each point on the 2D model depends on the two axis “*X* ” and “*Y* ”.

The impact factors (IF) used to develop the GASNV are just examples that show how the information taken from the 2D model can be useful to the optimization method execution. Four examples of IF were used, and each one of them received a value that is used to calculate an individual fitness. The IF’s and the values used were: unpaved (0.5); paved (1.0); public utility (for example, hospital areas, 1.5) and historical preservation (for example, cobblestone street, 2.0). These values and IF’s are totally didactic and a deeper study can be made in order to add different information seeking adaptation to specific situations as well as better results.

Just to give a better idea of how the fitness is calculated and taking as an example the individual shown on the Figure 4, the distance between “*P*1” and “*P*2” (assuming that the IF of this connection is unpaved) would result in an equation like the one showed on 3 and after defining the distance value, the fitness of the refereed connection would be as shown on 4.

$$Distance = ((x2 - x1)^2 + (y2 - y1)^2 \quad (2)$$

$$Distance = ((1 - 0)^2 + (0 - 0)^2 \quad (3)$$

$$Fitness\ Function = \frac{1}{1 * 0.5} = 0.5 \quad (4)$$

4.2.3 Selection and Reinsertion Methods

The roulette selection method has been chosen to select individuals for crossover. In this method, each individual has a piece of the roulette which is proportional to its fitness value. After mounting the roulette a random value is obtained and the matching individual in the roulette is used in the crossover process. When two equal individuals are drawn the system automatically generates another random value, until the moment that two different ones are chosen. In order to maintain the best chromosome (or group of better chromosomes) when defining the new population, GASNV was created using elitism method. Elitism can very rapidly increase performance of GA, because it prevents losing the best found solution.

4.2.4 Crossover

A hybrid of the Partially Matched Crossover (PMX) was created to suit the sewer network problem. PMX normal method follows four main steps. Two chromosomes are aligned. Two points are randomly selected along the strings, defining a matching section. The matching section is used to execute the gene reorganization. Alleles are moved to their new positions in the offspring. This process is shown on Figure 6.

The modified PMX crossover used on GASNV begins its execution by checking the compatibility of the selected individuals, because when they have no common points the crossover becomes useless. After testing the compatibility a random cut point is chosen on the first individual, and if the second one has this same point (coordinates x and y) the genes positioned after the cut point on the second individual are switched with the ones of the first individual. This process creates better solutions when the switched part has a better fitness, and by using the compatibility verifying process no useless crossover is executed. Figure 7 shows an outline of how the crossover is executed on GASNV.

4.2.5 Mutation

It is important to clarify that the GASNV does not use the mutation operator. Like explained before, the initial population created on the GA has just valid individuals and all the executed crossover considers the compatibility between individuals. In other words, all the individuals that are created on the initial population as well as the ones created by the crossover operation are valid individuals that can be used as a path to install the main pipe of the sewer network system. Normally the mutation method provides an important genetic variation for the system, creating genes that would not be found in other way on the population, but in this problem case, because of the problem characteristic and how the GA was created, the genetic variation provided by the mutation operator has been considered unnecessary.

4.3 Visualization Models

After the GA execution, a 3D model is created aiming at a better visualization of all the provided information. In this model the user has the possibility to navigate on a “*virtual city*” to see in a more intuitive way the path that will be used to implant the main pipe of the sewer network. To navigate on this model the user must use the arrow keys “*up*”, “*down* ”, “*right* ” and “*left* ”. The

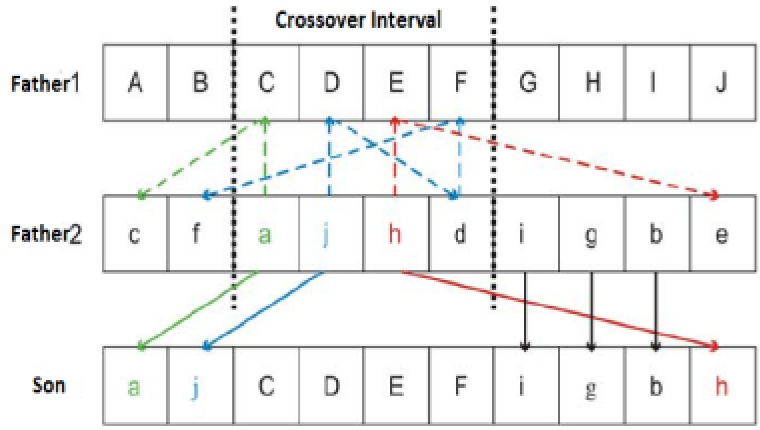

Figure 6. Partially Mapped Crossover

navigation process was created using two geometric transformations (glRotate and glTranslate). Areal world map was used to create the 2D model (using OpenGL), so the information obtained by the system execution is not merely didactic. The 3D model was created based on the 2D one and some fictitious buildings were used to simulate a real world situation. All the 3D models were designed using 3D Studio Max.

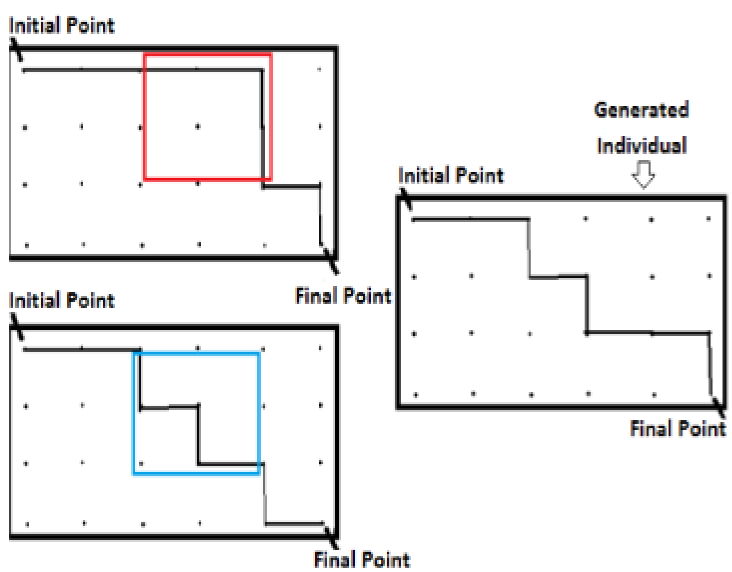

Figure 7. Crossover Outline

5.4 Statistics

During the development of the system some statistics were necessary to create graphics showing project estimates. The first data collected was an average of how many people exists in the 2D model. On GASNV every meter of pipe in the network corresponds to 2.5 people on the region. Based on this data the system calculates how many meters of pipes exists at the network created by the GA and multiplies this value for 2.5. The result is an estimated number of people in the network area.

The flow necessary to implant the network is very important to the system execution. In this paper the equation 6 was used to estimate the flow needed on the network. Each of this equation's variables is explained ahead.

Variable "*p*": represents the population on the network area. The value estimate to this variable is obtained as explained at the beginning of this section.

Variable “*q*”: obtained by the equation 5 and represents an estimate value of consumption.

Variable “*Cr*”: represents the amount of water which, after entering the residence by the water network, returns to the sewer network system. In this paper the value used for “*Cr*” was 80%.

The last formula used in the system is related to the nominal diameter (ND) of the network pipes. The equation used to establish a value to a pipe’s ND was 7. At this equation two variables were necessary: “*f* ” represents the flow of a network area and “*r*” means radius.

$$q = \frac{Population}{100} \tag{5}$$

$$Q = (p * q) + \frac{Cr}{1000} \tag{6}$$

$$ND = \sqrt{\frac{f}{2r}} \tag{7}$$

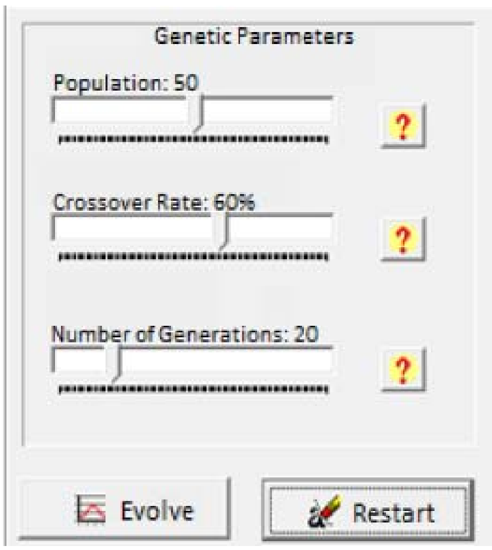

Figure 8. Genetic Parameters

5. System Operation and Results

The system requires three parameters to be executed: population size, crossover rate and number of generations (first, second and third boxes respectively on Figure 8).

After defining the three parameters the user must choose two points on the 2D model to be used as initial and final points of the main pipe of the network. The Figure 9 shows how the 2D model is exhibited, and the red points all over the map are the coordinates that can be selected by the user (to represent the initial and final points) before initiating the GA processing. In this image the initial and final defined points are inside yellow squares. The parameters were also defined but in order to have a better visualization of the map this part has been omitted. The parameters can be observed on Figure 8.

The different design used in the streets represents the type of impact factor (IF) of each one and in the top-left part of Figure 9 the labels explain all the IF’s. The first one (in blue) represents that the sewer network will begin in that area. The second one (brown) is used to define that every street in brown color represents the unpaved IF. The third label (gray) defines that every street in gray colorrepresents the historical preservation IF. The fourth one (black) represents the paved IF and finally, the white and red labelrepresents that the streets with these colors will receive the public utility IF value during the execution of the system.

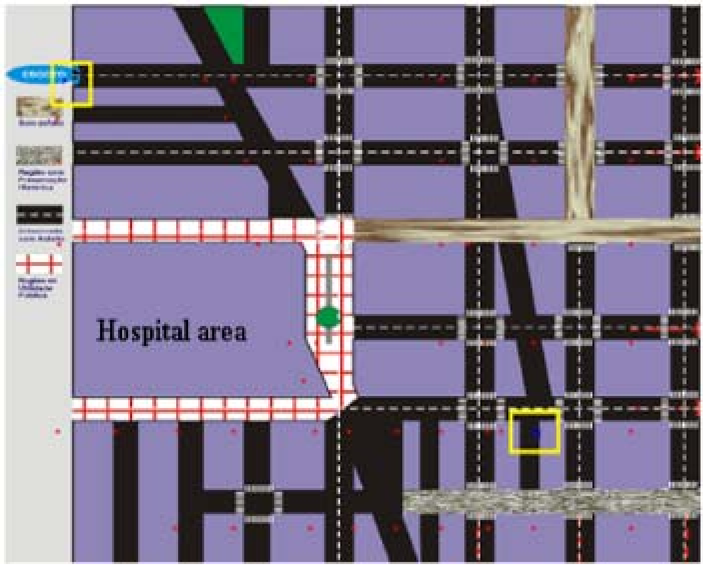

Figure 9. GASNV 2D interface with initial and final points selected

After establishing the genetic parameters and the initial and final points the user can initiate the GA processing. The proposed sewer network is shown on the 2D model with the main pipe represented by a green path and the auxiliary mesh in blue. The Figure 10 shows an example of a proposed network, using the parameters shown on the Figure 8 and the initial and final points defined on Figure 9.

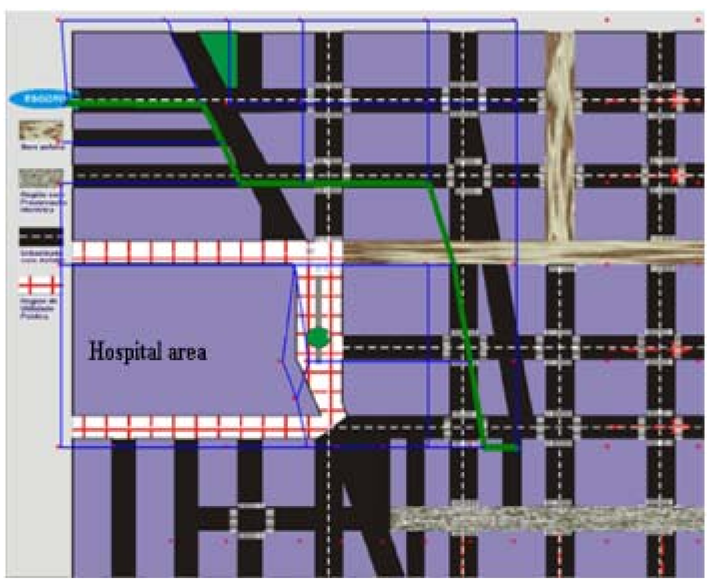

Figure 10. Example of sewer network proposed by GASNV

The proposed network has a group of estimates given by GASNV. This data (flow values, pipes length, diameter, excavation price and pipes price) is displayed on a proper interface. Figure 11 shows how the estimates are presented onthe interface. It is important to say that the 2D model has also an area where the most relevant estimates are displayed to the user. This area is positioned at the low-right part of the model, under the "*Genetic Parameters*" area, like shown on Figure 12.At the referred area, the user has access to the most relevant data involving the project, like initial flow, main pipe length,auxiliary network pipes length, final flow and the estimated price.

The last interface displayed by the system is the 3D one. This interface is responsible for allowing the user to navigate in a virtual city and thereby observe the proposed network in a more intuitive way. The data used to ”build” the referred interface is taken from the 2D model, and the buildings showed on it are didactic - they are used just to exemplify the possibilities given by GASNV. The Figure 13 shows how the virtual city is presented to the user.

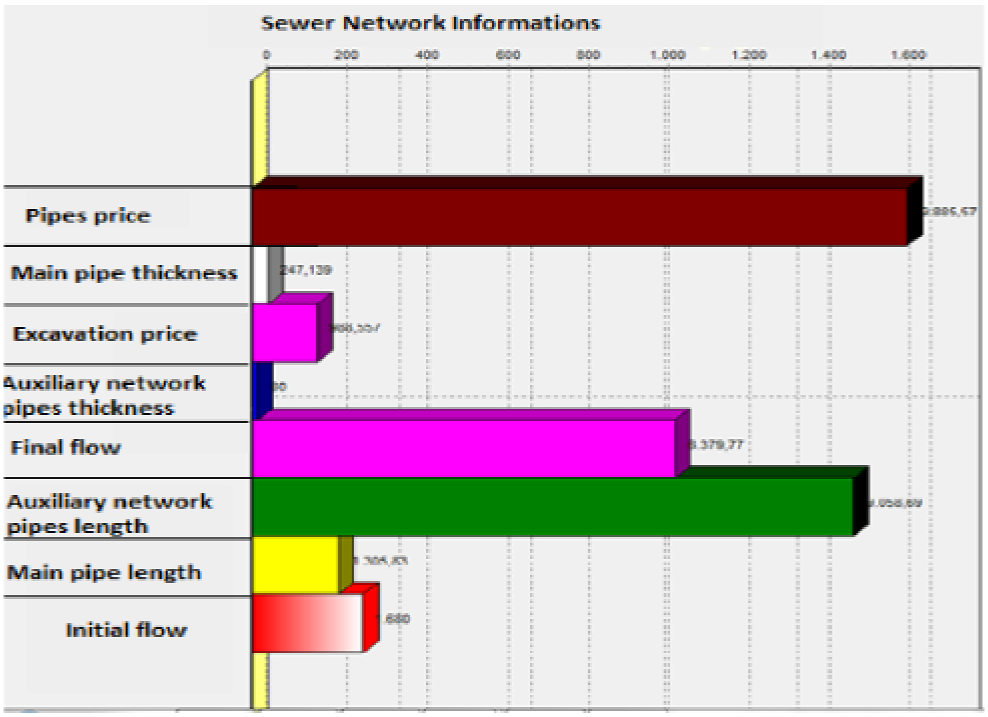

Figure 11. Interface responsible for display the project data

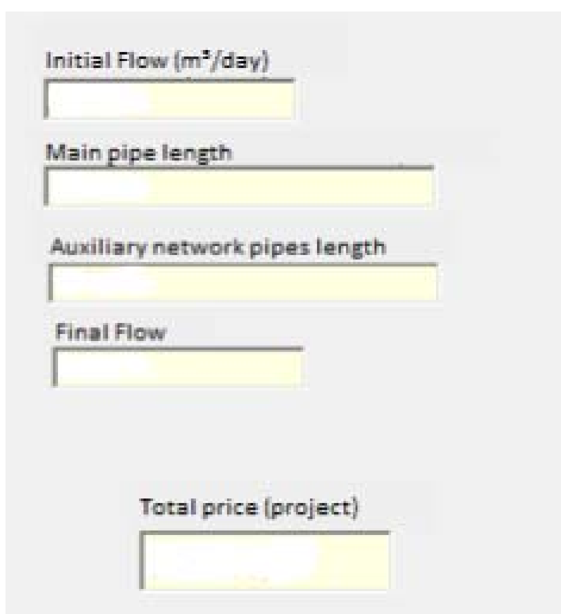

Figure 12. 2D model area where the main estimates involving the project are displayed

The tests performed in the system were always satisfactory. The genetic algorithm used as parameters the following numbers:

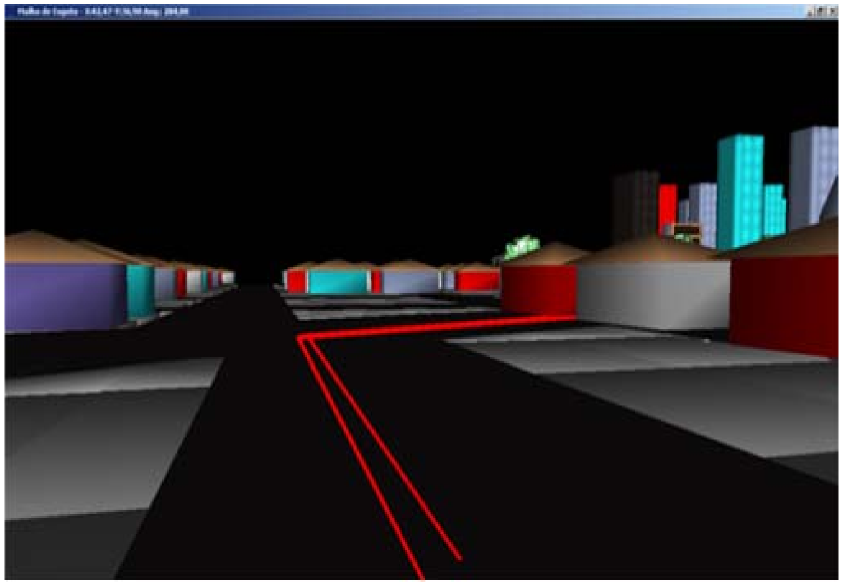

Figure 13. GASNV 3D interface

population size of 50 individuals, crossover rate of 60% and 20 generations. A lot of pairs points were tested as initial and final, and for all of them, the GA was capable of proposing a sewer network. These tests corroborate with the results obtained on the study made by [1] and reaffirm the great capacity of the GA method to deal with piped sewer networks optimization.

The combination of GA and IV methods showed good results, considering that the data provided by the optimization method became much simpler to analyze and that a lot of estimated data were inserted in the interface, turning the user experience more profitable. Some advantages arise specifically because of the 3D model, which provides a more intuitive and realistic analysis.

The initiative of creating a system combining an admittedly effective optimization method (Genetic Algorithm) with a good Information Visualization technique proved to be a valid experience, providing different ways to analyze the data andallowing researchers from other areas (not from computation) to have an easier assimilation of the information, therebymaximizing the results obtained by the system. GASNV system may also be considered as a tool that can be adapted and used to mitigate the great difficulties faced by some countries at the sanitation area.

6. Final Remarks

As discussed previously on this work, the whole world suffers from problems related to sanitation. The referred problems are very heterogeneous but in general, the cost and the lack of methods to help taking decisions related to them are the most important variables to be considered.

Based on the sanitation difficulties faced all over the world, this work proposed a method to support the optimal decision related to the implantation of piped sewer systems. The system proved its capacity to deal with the great amount of variables involved in the problem and the optimization method showed no limitations. GASNV is a first version yet, so, a lot of improvements can be made on the genetic algorithm model and on the visualization models. The networks used by [1] to test the genetic algorithm can be used in order to better evaluate the system, always seeking its improvement.

As a final remark it's important to stress the purpose of creating a system that allies an admittedly good optimizationmethod (GA) with improved Information Visualization methods, providing a system that has a less complex way of displaying the proposed information.

7. Future Works

In future works the following adjustments will be made: Create the Visualization models based on the networks used by [1] in order to have a better estimation of the system potential. Improve the Genetic Algorithm model. Change the selection method from roulette to stochastic tournament and evaluate the benefits that this method can bring to GASNV system. Use the mutation operator and evaluate the possible improvement brought by it.

Conduct a detailed study in order to find other variables that may be optimized by the system.

Conduct a detailed research in order to improve the visualization models, based on the profile of GASNV users.

8. Acknowledgement

Authors acknowledge the Brazilian research agency FAPEG for its financial support.

References

[1] Afshar, M. H. (2006). Application of a Genetic Algorithm to Storm Sewer Network Optimization, *Scientia Iranica*, 13 (3) 234-244 (Sharif University of Technology).

[2] David E. Goldberg. (1989). Genetic Algorithms in Search, Optimization and Machine Learning, Adison-Wesley.

[3] KIRNER, C., KIRNER, T. G. (2007). Simulation and Modeling: Current Technologies and Applications., Idea Group, 391-419.

[4] UNICEF. (2010). Progress on Sanitation and Drinking-water, World Health Organization.

[5] Nathan Yau. (2010). Visualize This: The FlowingData Guide to Design, Visualization, and Statistics, Indianapolis: Wiley Publishing.