Genetic Algorithm Optimization for Water Supply Systems Planning

Zohra Bakkoury and Driss Ouazar
Ecole Mohammadia d’Ingénieurs, Mohamed V University, Agdal, Rabat, Morocco
bakkoury@emi.ac.ma

Abstract

Water supply systems are built ahead of time to meet future demand. Investments are made in steps to cover needs up to a planning horizon. The objective of long term planning optimization is to come up with the least cost scenario for investment decisions. Potential savings could be very significant and influence the decision to invest or not for a given project. These also allow the choice of a feasible and optimal alternative amongst several project scenarios.

The present paper describes new approaches to long-term planning optimization using genetic algorithm search methods. The new approaches consist in determining the best phasing for long-term water projects. The formulation of the optimization problem is described along with three different genetic algorithm approaches. A case study based on the long-term water supply requirements of a region in North America is investigated. Finally, the results for the different GA formulations are presented and discussed.

Keywords: genetic algorithms, water systems, planning, optimization

1. Introduction

Long-term capacity planning aims at providing capacity in water systems beyond that needed to satisfy immediate requirements. Facilities are usually sized with sufficient capacity to meet anticipated flows several years into the future. The amount of excess capacity to be provided is a function of the conflict between economies of scale and the time value of money reflected by discount rates. On the one hand, economies of scale make it attractive to build beyond immediate needs as incremental costs are often small; on the other hand, decision-makers will tend to postpone investments in facilities that remain unproductive for long periods of time, as this can result in considerable savings in the present value of total investment costs. A proper planning process should take these two factors into consideration while satisfying demand forecasts over a planning horizon.

The few models proposed by researchers so far for optimal scheduling have been based on a number of assumptions which make them inapplicable to realistic systems. In current practice, phasing of water supply projects is still widely based on a trial method in which a set of solutions is examined and then changes made in some of them until a satisfactory solution is obtained. A low cost feasible solution is kept without any criterion of optimality. The solution is further and further far from the actual optimum when the size of the project becomes significant. There is therefore scope for achieving project phasing much more efficiently and consistently by using an optimization technique rather than the traditional trial-and-error method.

Most research work in the optimization of water systems design has considered only the static process, [1], [2], [3], [4], [5], [6], [7], [8], with little research dedicated to long-term planning. [9] expressed the optimal design period in years as a function of discount rates and scale factors. [10] proposed an approach to long-term optimal planning for tree networks. [11] considered a formulation under imprecise data while [12] presented a model that jointly optimizes pricing and
capacity expansion decisions for a public water supply system. [13] developed a multi-horizon optimization procedure to find the best way to invest progressively a limited amount of money in the rehabilitation of a water distribution system. [14] proposed an approach based on dynamic programming using the property of consecutive groups.

Costs of water facilities are very difficult to estimate and the use of inappropriate cost functions may bring inaccuracy to the models. Cost functions are in reality discrete. Because of the arbitrary character of these functions, mathematical approximations are difficult and can lead to entirely wrong estimations.

As it is impossible to define one smooth cost function for water supply project capacity planning, and as the costs include several types such as initial capital cost, incremental cost and running cost, it is usually not possible to solve the optimization problem in a conventional mathematical way. The number of components including pipelines, intakes, treatment plants and pumping stations, and their interaction, give a high degree of complexity to the problem. The number of alternatives is also very high and makes it almost impossible to evaluate all of them.

The present paper describes new approaches to long-term planning optimization using genetic algorithm search methods. The new approaches consist in determining the best phasing for long-term water projects. The formulation of the optimization problem is described along with three different genetic algorithm approaches. A case study based on the long-term water supply requirements of a region in North America is investigated. Finally, the results for the different GA formulations are presented and discussed.

2. The Long Term Planning Problem

The planning work in a water supply project is undertaken to determine the details of phasing related to the implementation of the physical system and its expansion over a time period. The challenge to project planners in the water supply sector is to meet growing demand for water as efficiently and economically as possible.

The objective of long term planning optimization is to come up with the least cost scenario for investment planning. Potential savings could be very significant, influencing the decision to invest or not for a given project. These also facilitate the choice of a feasible and optimal alternative among several project scenarios.

To minimize the present cost of all components necessary to satisfy growing demand over the project horizon, components can be installed in several stages while satisfying the growing needs of water. Determination of the scale of each stage and its timing depends upon the demand projection, since the main objective in water supply projects is to meet the demand. The size of the project also depends upon the relevant discount rate and the economy of scale factor of the works involved in the project.

Long-term capacity planning optimization aims to determine the timing and duration of each stage in order to minimize the total present cost of the project.

[9] considered equally sized expansions by extending an earlier model developed by [15] and [16] to determine the optimal design period expressed in years as a function of discount rates and scale factors. Three models were presented assuming demand to grow linearly over time: an expansion model, a model with an initial deficit, and a waiting period model.

The work of Lauria et al. was based on a number of questionable assumptions. Questions of applicability may be raised about linearly increasing demand. The assumption that cost functions
for both initial construction and expansions are identical and constant over time, as stated by the authors themselves, is not well founded. The assumption of an infinite planning horizon also seems to be unrealistic and could have been avoided. Optimality conditions could have been largely reinforced by integrating operating costs. Finally, determining the optimal design periods depends on the accuracy of economy scale factors of waterworks as a whole system, which are difficult to estimate and vary from one country to another.

[10] presented a model for long-term optimization of tree water networks based on linear programming. The network is represented by a tree whose branches coincide with the layout of the pipes and whose nodes are the consumption points and the junctions of two or more branches. A branch is defined as being composed of a group of sections, each one being associated with a different pipe diameter. The network contains a single source corresponding to the origin node of the tree. Demand at each node is known throughout the planning period. The problem variables represent section lengths to be installed in each branch as well as heads of the pumps in the network and head-drops of the PRV. The locations of existing and future pumps and PRV in the network were assumed to be known. The objective function consisted of minimizing the sum of present worth values of investment costs and operation costs over the planning horizon. Investment cost was composed of costs of pipes, pumps and valves. Operation costs included the operation costs of pumps.

The optimization over time was formulated as a multi-stage linear program. Each stage corresponds to the state of the network during a specified interval of the planning period. The linear model used the revised simplex algorithm to solve the problem of optimization.

The main limitation of this model resides in the assumption of a tree network with a single source node, a topology which is over-restrictive. Approximating the costs of pipes, pumps and valves by linear functions makes the model inapplicable to real life systems. In addition, treatment works and reservoirs have not been considered by the model.

[11] studied expansion planning of water supply and wastewater treatment systems based on a linear programming formulation with imprecise data. The problem was formulated as a stochastic multi-criteria decision problem, taking into account cost, reliability and environmental protection criteria. The constraints consisted of maximum water source capacities and water demand by particular users.

As in [9], this model considered cost functions to remain constant over the planning horizon. It considered equal length expansion periods, which is unlikely to lead to the overall optimum in practice. Also, the definition of the cost term relating to environmental quality lacks precision.

Given the importance of price in water systems [12] presented a model which jointly optimizes pricing and capacity expansion decisions of a public water supply system. The model aimed at identifying the water price horizon so as to maximize the present value of net benefits minus resource costs over the planning period. Constraints on the range of water price, rate of price change, and financial cost recovery were included in the formulation. Demand per year was expressed as a function of price and included losses in the system. The general model was solved using a discrete dynamic programming algorithm. Decision variables were the optimum price and the capacity increment of the system. The capital costs were based on a function of scale economies while operating and maintenance costs were considered constant up to the system capacity and infinite thereafter.

The model assumed a single price per unit of water for all users, which may not be acceptable to water pricing authorities, especially for systems where users have different profiles, i.e. different consumption, activities, or geographic location. Moreover, cost functions used in the model may be inaccurate.
[13] developed a multi-horizon optimization procedure to find the best way to invest progressively a limited amount of money in the rehabilitation of a water distribution system, according to different horizon funds. The method is based on a multi-objective optimization procedure using a Structured Messy GA (SMGA). It involves finding a set of complementary solutions producing the maximum benefit to the system, where each solution is a part of the final solution.

[14] presented a generalization of Manne’s model based on the following assumptions:
- Demand grows non linearly over time;
- The equipment has an infinite economic life; and
- Whenever demand catches up with the existing capacity, x units of new capacity are installed.

The installation costs that result from a single capacity increment of size x are assumed to be given by a cost relationship in the form of a power function $kx^\alpha$ ($k>0, 0<\alpha<1$).

As the time for subsequent calculations, any such point as $t_0, t_1, t_2$ represents a time at which the previously existing excess capacity has just been wiped out. Such a point is known as a 'point of regeneration'.

A dynamic programming technique using the property of consecutive groups was applied. The total cost of expanding $N$ groups included the discounted installation cost incurred directly at the end of the current cycle and the sum of all installation costs incurred in subsequent cycles.

The cost function in this model lacks accuracy and the assumption of infinite economic life for equipment is not well justified.

3. Optimization Problem Formulation

The optimization problem consists of finding the best size, location and timing within the planning period of investments to increase system capacity. The water network is represented as a set of links. Each link connects two nodes and is composed of a pipeline and other elements such as pumps, reservoirs and treatment plants. The objective function consists of the present worth of the investment costs at each stage as well as the present worth of operation and maintenance costs of all the network elements.

The general optimization problem for capacity planning in a water distribution system can then be formulated as follows:

**Objective** Minimize total present worth of costs

**While** Satisfying a time varying demand over the project duration

**Subject to** Constraints

3.1. Variables

The decision variables define the capacity increments and timing for each stage until the end of the planning horizon. Each link increment value will vary between specific upper and lower bounds that depend on the link minimum and maximum capacities.

The optimization variables are discrete, i.e. they are allowed to take any value from a discrete set of values distributed within specified limits.
A decision corresponding to a given link at a given year represents a possible action on this link, i.e. an increment value to add to the link capacity in the specified year, or a “do nothing” option when the capacity of the link is not changed during that year.

Typically, the number of optimization problem decision variables will be equal to \( N_{\text{links}} \times N_{\text{years}} \), where \( N_{\text{links}} \) and \( N_{\text{years}} \) are the number of links in the network and the project duration respectively.

3.2. Constraints

The operational and design limits of water systems impose a certain number of constraints on the network. These constraints generally include:

- Minimum and maximum capacity constraints.
- Capacity increments should have positive values, i.e. system capacities of different components are generally supposed to grow or remain constant.
- Total capacity of the water system should be greater than the peak demand; systems with deficit are considered as infeasible.

Other constraints may be considered depending on system characteristics and the functionalities sought from its operation.

3.3. Objective Function

The objective function to be minimized is composed of cost functions including capital, operating and maintenance costs for all components in the water system.

\[
\text{Min} \left( \sum_{j=1}^{N} \frac{(I_j + O_j)}{(1+r)^j} \right)
\]

(1)

Where

- \( I_j \) represents the sum of capital investments for year \( j \)
- \( O_j \) total operating and maintenance costs for year \( j \)
- \( r \) discount rate
- \( N \) the duration of the planning horizon

Capital cost, i.e. initial capital cost and incremental cost, consists of investment costs of pipes, pumps, valves, treatment plants and reservoirs. Operating or running costs include mainly energy costs for pumping and treatment costs.

4. Genetic Algorithm Approaches

Genetic algorithms (GA or GAs) are a part of evolutionary computing that derives from Darwinian evolution. They were originally introduced by John Holland in the 1970’s at the University of Michigan, and have since then been further developed by [17], Goldberg [18] and other researchers. These algorithms are best suited to solving combinatorial optimization problems, which cannot be solved by using more conventional methods.

GAs have been successfully applied to a number of water system optimization problems, such as design, operation, calibration and rehabilitation. Over the last 10 years, the number of studies
related to the use of genetic algorithm optimization in water supply and distribution problems has considerably increased.

[19] presented an interesting review of GAs application to water systems, including layout and pipe sizing problems, rehabilitation, pressure regulation, calibration and pump scheduling.

Three GAs formulations are investigated in this paper and applied to a real life case study:

- A simple GA (SGA)
- A fixed length GA (FLGA)
- A messy GA based on the SMGA developed by [20]

An integer encoding was adopted for the three GA formulations. The simple GA encodes all possible variables in a long string, whereas the FLGA and SMGA omit genes which represent inactive variables. For the phasing problem, an active decision variable corresponds to a year and a link where the capacity is increased. Inactive decisions, then, represent links and years where the capacity remains unchanged (variables corresponding to a “do nothing” option).

Encoding examples for the three genetic algorithm optimization approaches are given in the following sections.

4.1. SGA Approach

The simple genetic algorithm used in this study involves the use of steady-state replacement, the two-point crossover operator, the uniform mutation operator, and the tournament parent selection scheme. More details about GAs can be found in [21] and [22].

The string length is equal to the number of links in the network multiplied by the number of years in the project duration. The integer value in each string position defines the increment value to be added to the corresponding link and year. The gene locus (position) in the SGA defines the year and the link while the allele (value) defines the decision.

Figure 1 is an example of a chromosome composed of 8 genes. The chromosome represents two links to be developed over four years. The gene $a_{i,j}$ represents the decision corresponding to link $i$ in year $j$.

```
  a1,1  a2,1  a1,2  a2,2  a1,3  a2,3  a1,4  a2,4
```

Figure 1: Capacity planning chromosome example for the SGA

4.2. FLGA Approach

Within the SGA approach, when dealing with real-life water systems, chromosomes representing feasible solutions for the phasing problem contain a considerable number of inactive decision variables. A decision variable is considered inactive when it represents a “do nothing” option. The fixed length GA used in this study tries to cope with this problem by limiting the chromosome length to a fixed number of genes. Hence, the string is set to a constant length during a GA run for the FLGA. The missing genes are then considered inactive.

Figure 2 shows an example of FLGA encoding where the string is split into three sub-strings for the phasing problem:

- The link sub-string: defining the network link references
- The decision sub-string: defining the decisions for corresponding links
- The time sub-string: defining the years in which the decisions take place
Link sub-string: 2 1 3 1 4 2 2
Decision sub-string: 3 0 5 1 2 3 4
Year sub-string: 4 2 3 2 1 1 4

**Figure 2: FLGA and SMGA encoding**

A gene is represented by three integers with the same position in the three sub-strings. In the example above, the first gene (2,3,4) indicates that for link number 2, in year 4, the decision is represented by integer 3. The decision allows coding a capacity increment depending on the number of possible alternatives for each link.

The FLGA developed in this study involves a steady state replacement and tournament selection method. One-point crossover is applied to the three sub-strings simultaneously at the same cut point. The mutation operator changes the three integers representing a single gene with a valid gene depending on the problem.

### 4.3. SMGA Approach

The SMGA used in this study is based on the structured messy GA developed by [20]. It is a GA which incorporates some of the principles of messy genetic algorithms [23], such as the type of coding and the variable length chromosomes. Essentially, the SMGA approaches the problem and builds up the final solution progressively, in a structured manner. Only active decision variables are included in a string of a small maximum length. The algorithm starts the process by considering one-element strings corresponding to a single decision variable from which it develops progressively longer and longer strings throughout the process. In the basic iteration, the best single elements are added onto the current population members to form a new population with longer strings, this being known as concatenation. More details about the SMGA can be found in [20].

However, some alterations to the SMGA defined by Halhal in [20] were necessary to adapt the algorithm to the phasing problem. The string is split into three sub-strings using the same coding as the FLGA, as shown in Figure 2, while the chromosome length varies between 1 and the maximum allowed length. Crossover and mutation operators are also changed as described in the FLGA section. On the other hand, gene duplication or over-specification is allowed in the present formulation. The first come first served rule, with a scan from left to right, is then used when two or several conflicting genes are present in the same chromosome.

For the three GA formulations, the same objective and fitness functions are used. Each chromosome represents a potential solution to the phasing problem and is evaluated to obtain its fitness.

### 4.4. Objective Function

When applying genetic algorithm optimization to long-term planning, the objective function on which the optimization search is based is the minimum present cost. The objective is thus to find the chromosome that has the minimum value of Equation (1). A penalty cost is added to the objective function when the constraints are violated, i.e. when the solution presents a deficit in satisfying water demand or the maximum and minimum capacity bounds are not respected. Therefore, the GA assigns a penalty cost for each solution if the network presents a deficit in satisfying the demand at any time over the project horizon. Estimation of the deficit requires a maximum flow algorithm to be run [24] to determine the actual network capacity. The deficit, if any exists, is then multiplied by a penalty factor ($k=25$ for this study). A penalty is also assigned
to each solution where the minimum and maximum capacity constraints are violated. The excess in capacity is multiplied by the same penalty factor $k$.

The total network cost of each individual in the population is then taken as the sum of the network cost plus the penalty cost.

4.5. Fitness Function

The GA computes the fitness for each solution in the population using the inverse of the total network cost, including penalty, such that the lowest cost strings represent the fittest individuals. The following formula is used:

$$Fitness_i = \frac{1}{total\_cost_i} \quad \text{for strings } i = 1, ..., Pop\_size$$

(2)

Where $Fitness_i$ represents the fitness of individual $i$ and $total\_cost_i$ the total network cost of solution $i$. $Pop\_size$, the population size.

4.6. Maximum Flow Algorithm

Any solution to the problem of long-term planning must be able to convey the maximum day demands from sources to demand nodes for every year in the planning horizon. To guarantee that demand is satisfied every year at each node, a deficit calculation procedure is implemented using the maximum flow algorithm [24].

The algorithm was developed by Ford and Fulkerson in 1956. Its objective is to find the maximum flow that can be carried over a network. The maximum flow algorithm generates a flow of maximum value from a node called the source, to a node called the sink. The flow in each arc or link is limited to the arc capacity. At each iteration, the algorithm attempts to find a path from the source to the sink in the residual graph. The capacity of an arc in the residual graph is the original arc capacity minus the cumulative flow assigned to this arc in previous iterations. If a path from source to sink is found, a flow of maximum possible value is sent through the path. The algorithm stops when no more such augmenting paths can be found.

If this maximum flow is less than the peak demand, then the solution being evaluated presents a deficit.

To apply the maximum flow algorithm to a water network, all sources are connected to a dummy node representing the single source of flows entering the system, while all demand nodes are connected to a dummy node representing the single sink.

5. Case Study: Long-Term Water Supply for a Region in North America

When the study was conducted, sources of water for the study region were only just sufficient to meet projected water needs up to year 2000. New resource development was needed to respond to increased usage and population growth up to 2031. Several possible new sources were identified. Only competitive sources in terms of cost were kept for the optimization process.

The forecast demand for water for each year to 2031 was provided for the different geographical centers of demand in the region.

The aim of the study was to determine the least cost combination of sources to develop for the case study region together with the optimum scheduling of their implementation.
5.1. Network Description

The system consists of a number of possible links connecting four possible sources of water (E, B, D, T) to four points of aggregated demand: W, X, Y, Z. A preliminary study based on cost criterion had previously identified several non-competitive sources, which were not included in the optimization process. Only viable schemes have been retained. The network remaining to be optimized is shown in Figure 3. Seven schematic links were proposed to supply the demand nodes from the potential sources. Arrows indicate the possible water transfer directions.

Measures for reducing consumer demand were also considered in the problem formulation, with appropriate costs. The optimization must satisfy the range constraints including capacity constraints and a zero deficit requirement, taking demand reduction measures into account when appropriate.

5.2. Optimization Problem Formulation

The number of variables or genes is equal to \( N_{\text{links}} \times N_{\text{years}} \), where:

- \( N_{\text{links}} = 8 \) including the 7 proposed links plus a dummy link to represent measures for demand reduction;
- \( N_{\text{years}} = 32 \), the project horizon.

For the SGA, the size of the chromosome is therefore \( 8 \times 32 = 256 \).

The optimization problem corresponding to the case study is then formulated as follows:

\[
\text{Min} \sum_{y=1}^{N_{\text{years}}} \text{Present}_\text{-Cost(network)}_y = \text{Min} \sum_{y=1}^{N_{\text{years}}} \sum_{i=1}^{N_{\text{links}}} \text{Present}_\text{-Cost(link)}_i_y
\]

Subject to

\[
\text{min}\_\text{capacity}_i \leq \text{increment}_i \leq \text{max}\_\text{capacity}, \forall i \in [1, N_{\text{links}}]
\]
\[
\min \text{ capacity}_i \leq \text{ current capacity}_i \leq \text{ final capacity}, \forall \ i \in [1, N\text{links}] \\
\text{deficit}_y = 0 \forall \ y \in [1, N\text{years}]
\]

(5)  

(6)

Where \( \text{present cost}(\text{network})_y \) represents the network present cost for year \( y \) including running costs and investments decided on year \( y \).

Total cost includes initial or incremental cost and running cost for all link components. Running costs for each link depend on the current capacity of the link. Cost estimations were developed for the construction of new pipelines, pumping stations, ground reservoirs, elevated tanks and treatment plants. These costs are based on the actual historical cost of similar facilities.

Hence, costs were provided for each link for a range of final capacities in steps of 5 or 10 MGD (Million Imperial Gallons per day) up to the maximum capacity allowed or envisaged. For each final capacity, the capital and running costs were detailed for the main components of each link (intake, treatment works, pumping stations, pipelines and storage) based on an incrementally staged development. Data specifications assumed that intakes and pipelines were not staged, except where the final pipeline capacity was to exceed 50 MGD, in which case a two-stage development of the pipelines was considered. The other components were based on development in 10 MGD steps, with an initial 20 MGD step where appropriate. The lake supply (D) had additional data for a 5 MGD scheme.

The capacity increments of each link will vary between maximum and minimum values and final capacities will also be bounded. Table 1 lists the initial and maximum capacities allowed for each link of the network along with the number of possible alternative actions. Given the individual options specified above, the total number of possible combinations for the 32-year period is calculated as approximately equal to \( 4.5 \times 10^{215} \). Although most of these combinations would represent impractical solutions, there remain a very large number of feasible solutions to choose from.

<table>
<thead>
<tr>
<th>Link reference</th>
<th>Initial Capacity</th>
<th>Max Capacity</th>
<th>Number of alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>[BM]</td>
<td>0</td>
<td>120</td>
<td>13</td>
</tr>
<tr>
<td>[GD]</td>
<td>0</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>[MR]</td>
<td>0</td>
<td>70</td>
<td>13</td>
</tr>
<tr>
<td>[RN]</td>
<td>0</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>[VR]</td>
<td>0</td>
<td>70</td>
<td>13</td>
</tr>
<tr>
<td>[VE]</td>
<td>0</td>
<td>120</td>
<td>13</td>
</tr>
<tr>
<td>[TP]</td>
<td>70</td>
<td>103</td>
<td>4</td>
</tr>
<tr>
<td>Demand reduction</td>
<td>-</td>
<td>-</td>
<td>4</td>
</tr>
</tbody>
</table>

Capacities in MGD

Table 1: Capacities and number of alternatives
Table 2 shows the number of alternatives for link [GD] from lake D corresponding to the four potential decisions counting possible increments 5, 10 and 20 MGD plus a “do nothing” option coded with integer value 0.

<table>
<thead>
<tr>
<th>Decision (capacity increment)</th>
<th>Integer coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 MGD</td>
<td>0 (do nothing)</td>
</tr>
<tr>
<td>5 MGD</td>
<td>1</td>
</tr>
<tr>
<td>10 MGD</td>
<td>2</td>
</tr>
<tr>
<td>20 MGD</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 2: Possible alternatives for link [GD]

Options for utilizing the existing supply T were provided as follows:

- Retain the existing maximum day capacity of 70 MGD
- Increase the maximum day capacity to 103 MGD in year 2000 at a capital cost of $103M
- Reduce the capacity to 45 MGD maximum day in year 2000 at a capital cost of $2.25M
- Reduce to zero in any year from 2000 onwards at a cost of $30M

Possible demand reduction measures were also specified, costs and effects being as follows:

- UFW (leakage reduction) reduces demand by 1.16 MGD over the whole region, and costs $354,000 initially, with $325,000 in the first year of operation, then $270,000 annually thereafter
- DSM (demand reduction) reduces demand by 3.08 MGD over the whole region, and costs 11.86M initially, with $170,000 annually thereafter

For this study, demand reduction was considered the same on both average and maximum days.

It was also assumed that any source could be implemented at any time from 2000 onwards, the capital expenditure could be considered concentrated in the year before a scheme comes on line, and that operational costs commence in the year the scheme comes on line.

The optimization problem was to determine the best combination of resources to develop and the optimal timetable for their staged implementation. The criterion used was the minimum present value in 1996 of capital plus operational costs from 1996 to 2031 based on a discount rate of 6% (9% interest rate with 3% inflation).

6. Capacity Planning Results

The three GA approaches SGA, FLGA and SMGA were applied to the long term water supply problem.

6.1. SGA Results

Table 3 presents a summary of the best relevant solutions obtained from several SGA runs using different seed numbers and control parameter values. Table 4 shows details of the implementation of the best solution, costing $627.3M.
<table>
<thead>
<tr>
<th>[BM]</th>
<th>[GD]</th>
<th>[MR]</th>
<th>[RN]</th>
<th>[VR]</th>
<th>[VE]</th>
<th>Existing [TP]</th>
<th>Demand Reduction</th>
<th>Cost $M At 6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>20</td>
<td>-</td>
<td>10</td>
<td>10</td>
<td>50</td>
<td>70</td>
<td>DSM UFW</td>
<td>654.4</td>
</tr>
<tr>
<td>40</td>
<td>20</td>
<td>-</td>
<td>50</td>
<td>70</td>
<td>DSM</td>
<td>-</td>
<td>664.4</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>10</td>
<td>50</td>
<td>70</td>
<td>DSM</td>
<td>-</td>
<td>627.3</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>50</td>
<td>DSM</td>
<td>UFW</td>
<td>632.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-</td>
<td>50</td>
<td>103</td>
<td>DSM</td>
<td>688.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Summary of the best solutions of the SGA

<table>
<thead>
<tr>
<th>Year</th>
<th>2000</th>
<th>2001</th>
<th>2004</th>
<th>2007</th>
<th>2009</th>
<th>2015</th>
<th>2022</th>
<th>2027</th>
<th>2031</th>
<th>Final Capacities</th>
</tr>
</thead>
<tbody>
<tr>
<td>[BM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>10</td>
<td>10</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td>[GD]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>[RN]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>[VR]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>[VE]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>10</td>
<td>10</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td>[TP]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td>DSM</td>
</tr>
</tbody>
</table>

Increments and final capacities in MGD

Table 4: Details of SGA best solution phasing ($627.3M)

The layout of the best solution found by the SGA is shown in Figure 4, having a present value at 6% discount rate of $627.3M. The solution involves early development of link [EV] from E to supplement the existing supply from T, which is retained at 70 MGD maximum day. The development of links [VR] and [RN] to supply Z from the west allows development of a lake supply through link [DG] to be postponed until 2009 when the source from E has reached its limit of 50 MGD. By 2015, Link [BM] is developed in three stages to reach 40 MGD capacity in 2027. Finally, demand reduction measure DSM is introduced in 2031 to avoid the need to develop further capacity for the last year of the planning horizon.
6.2. FLGA Results

The FLGA was investigated with the objective of limiting chromosome length, avoiding a large number of inactive variables and reducing the memory use.

To assess the influence of chromosome length on the algorithm performance, four optimization runs were completed using a different chromosome length for each run: 15, 20, 30 and 80 respectively. Each length was run three times with different seed numbers. Total number of evaluations was maintained at 500,000. The results presented in Figure 5 show that the best solutions for the FLGA were obtained with chromosome lengths 20 and 30. Moreover, when the
length increases, the algorithm slows down significantly. When increasing the chromosome length to code more than 85 genes, the individual using three sub-strings in the FLGA requires more memory space than the same chromosome with SGA encoding. The 85 fixed-length individual involves 255 integers to code 85 decision variables while the SGA uses 256 integers to code all the 256 decision variables of the problem.

Comparison with the SGA results show that fixing the chromosome length confines the exploration and exploitation of the search space potential of the GA. The solutions found are far from the previous optimum, although they still represent good solutions to the problem.

6.3. SMGA Results
The SMGA adapted to the phasing problem was successfully applied to the case study, identifying a solution costing $620.2M at 6% discount rate. Details of implementation of this solution are given in Table 5.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[BM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td>10</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>[GD]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>[RN]</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>[VR]</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>[VE]</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>[TP]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>Reduction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DSM</td>
<td></td>
</tr>
</tbody>
</table>

*The negative sign indicates a flow in direction [RV]

Table 5: Details of SMGA best solution phasing ($620.2 M)

Compared to the best solution found by the SGA, this solution involves the development of the same schemes with a different development schedule. The project manager can easily switch to the solution identified by the SMGA starting from year 2007, making a saving of more than $7M.

To assess the influence of using the best single-element strings in the concatenation process, comparative runs concatenating single elements at random were performed. Three different runs were completed using different seed numbers with each concatenating type. The concatenation step was maintained at one element. The results show that the two types of concatenation found the same optimal solution. However, this solution had been identified earlier when concatenating with the best single elements. This is possibly due to early chromosomes formed with the best elements accelerating the search process by contributing to forming (parts of) good solutions in early generations, leading through selection and crossover to the best solution. As the concatenating process goes on, the best single elements have more chance to be duplicated in the same chromosome and no improvement occurs in the final steps of the process.
On the contrary, when concatenating with random elements, the process of improving the solutions in the successive generations is slower at the beginning of the GA run.

6.4. Sensitivity Analysis
The first analysis investigates sensitivity of the solutions to changing discount rates. Three discount rates were examined. The results do not show any noticeable dependence on the interest rate, proving a good stability of the solutions identified. Figure 6 shows the different costs for different discount rates of the best solutions for the three GA methods.

![Figure 6: Best solutions at different discount rates](image)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[BM]</td>
<td></td>
<td></td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>[GD]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>[RN]</td>
<td>10</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>[VR]</td>
<td>10</td>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>[VE]</td>
<td>30</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>[TP]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>70</td>
<td></td>
</tr>
</tbody>
</table>

Reduction

Table 6: Best solution phasing for 90% Demand ($554.3M)
The second analysis investigates the sensitivity to changing demand forecasts at a constant discount rate of 6%. Examination of the solutions obtained for schemes based on demands of 90% and 70% of the full demand forecasts presented in Figure 7 shows that the different demand scenarios involve similar development strategies in the early years, making the best solutions found very robust with respect to unexpected reductions in future demand. This allows the project manager to react efficiently to possible changes in demand scenarios due to uncertainty in forecasts. Total savings could be very significant. Table 6 and Table 7 present the implementation details for the best solutions based on demands of 90% and 70%. The layout of these solutions is given in Figure 8 and Figure 9.
Figure 8: Best Solution for 90% Demand ($554.3M at 6%)

Figure 9: Best Solution for 70% Demand ($422.7M at 6%)

UFW in 2016

DSM and UFW in 2021
7. Conclusions

The approach to long-term planning optimization in water projects as formulated in this research involves:

- The use of discrete variables;
- Capacity increments for proposed links in different years of the planning horizon and demand reduction measures as the variables;
- The present worth for capital and operational costs as the objective function.

Feasible solutions must have no deficit in supplying demand and no excess capacities beyond maximum allowed values.

In this paper, three approaches based on different GA formulations have been investigated for long-term water supply. The three methods have been applied to a real-life case study to supply a major region in North America with water up to 2031. The best combination of water supply sources and their development schedules were identified. Several near optimal solutions have also been presented.

The comparison of results showed a good performance of the SMGA approach for the phasing problem with major reduction in memory use.

The SGA also proved to be efficient in finding near optimal solutions even with considerable memory use and computation time.

The FLGA was the least efficient of the three approaches, probably because the fixed length chromosomes limit the exploration/exploitation ability of the genetic algorithm when applied to the long-term planning problem.

The results also showed that the optimum solution is not sensitive to small changes in the discount rates used to establish present value of capital and operational costs. Examination of the solutions obtained for schemes based on demands of 90% and 70% of the full demand forecasts has shown that the different demand scenarios involve similar development strategies in the early years, making the best solutions found very robust with respect to unexpected reductions in future demand, allowing the project manager to react efficiently to possible changes in demand scenarios due to uncertainty in forecasts. Savings could be very significant.

In conclusion, genetic algorithms have been successfully applied to the problem of selection of the most economic sources to develop and the scheduling of their development. The results of GA optimization prove its high potential in identifying good solutions for the phasing problem.

Finally, the rapid re-run option enables the planner to perform effective sensitivity analysis to assess the impact of revising demand allocation or changing discount rates. Alternative cost scenarios may also be easily examined. In the water supply case study, further economies could result from refining the way in which flow–dependant operational costs are estimated.


Z. Bakkoury and D. Ouazar  Genetic Algorithm Optimization for Water Supply Systems Planning


