Efficient Aquifer Remediation Using Evolutionary Processes For Decision Making

M.F.Yeo, E.O.Agyei, University of Adelaide, South Australia
and D.Armstrong, Lisdon Associates, Adelaide, South Australia

Abstract The economic cleanup of aquifer contamination necessitates decisions as to the locations and pumping rates for the extraction wells. Even allowing for topographical and political restrictions as to the well locations, there are usually a large number of possible well locations together with a number of different economical pumping rates. In addition to these decision variables any pump when installed can be either pumping or else not pumping at different times during the cleanup operation. The number of possible solutions to a problem is therefore very large. For example, if there were ten possible well positions, four rates of pumping at each well including not pumping and four time intervals when the pumping rates can be changed there are 2^{80} ($=1.21\text{x}10^{24}$) different possible solutions. Due to the discontinuous nature of the problem, caused by the discrete pumping rates, linear optimisation techniques are inappropriate. Problems involving optimisation of discontinuous functions are now being solved by evolutionary optimisation schemes such as Genetic Algorithms (GAs). The process of GAs is discussed and then its use for solving the extractions problem is described. The optimisation is carried out to determine the minimum cost of the remedial action, when all costs are incorporated into the cost of pump installation and into the cost of each unit volume of polluted water extracted. The technique is applied to a theoretical problem of reducing pollutant levels in an aquifer to acceptable levels. The conclusion is that the GA optimisation will produce realistic solutions. There is no guarantee that the GA solution is the global optimum but it will be of acceptable efficiency that may be implemented or else used as a starting point for further investigation.

1. INTRODUCTION

When pollution is detected in an aquifer it is usually considered so unacceptable that its removal is almost always obligatory. Assuming that the extent of the pollution and the aquifer flow and contaminant transport coefficients are known then, the design of an acceptable, realistic and economic remediation scheme will involve decisions with input from politics, topography and engineering. A common mode of remediation is the pump and treat approach and this paper will limit considerations to this type of scheme. The locations of the pumps are limited by both political and topographical considerations. Even allowing for such restrictions there are usually a large number of possible pumping locations. The engineering decisions that then have to be made are which sites to use, what sizes of pumps to install, the pumping regime, the treatment process and the method of disposing of the treated water.

Pumps are designed to operate at specific rates for economic operation. The selection problem is therefore approached by selecting the pump sizes available and then determining where they should be located. All costs involved in the remedial process can then be included in two costs: the cost of installing a pump and the cost of extracting a unit volume of polluted water. The pump maintenance, running costs including depreciation, and pollutant treatment costs can be included in the unit extraction cost.

A decision also has to be made as to the level of pollutant that may be considered acceptable. It is uneconomic to remove pollutant that reduces the pollution below this acceptable level. The means of treatment and disposal of the pollutant must also be considered. For the example used in this paper the polluted water will be pumped to an evaporation pan and the resulting residue removed at a later date for treatment. However it would have been just as easy to have selected air stripping and reinjection of the treated water back into the aquifer or any other scheme.

Optimisation is necessary to minimise the total cost of the remediation scheme. Gradient based techniques are inappropriate for fixed pumping rates, as the resulting flow conditions vary in a step fashion when a pump is turned on or off or when a new pumping rate is introduced. These techniques usually require that the parameters can vary smoothly between limits to generate a sensitivity matrix. Recently, evolutionary based optimisation techniques have been used to solve non-linear and discontinuous problems as they do not utilise the derivative of the objective function to be optimised but use the function itself. They have been proved to be robust in overcoming the problems in optimising such problems.

Genetic algorithms (GAs), [Goldberg, 1989] are one of a group of evolutionary techniques, that are based on the process of natural selection that occurs in the genetics of a species. The algorithm was developed from the work of Holland (1975). Basically the variables involved in any problem are encoded as a string of binary digits, the 'genes' of the problem. Randomly selected solutions are compared for efficiency and the best are selected for further investigation. Pairs of the best solutions, or 'parents', are combined and their genes are resequenced to produce two new solutions, or 'children'. An occasional gene digit is switched to implement 'mutation'. The children then form the parents of the next generation and the process is repeated.

GAs have recently been used for a number of problems involving groundwater pollution, [McKinney and Lin, 1994; Ritzel and Eheart, 1994; Cieniawski et al, 1995]

2. GENETIC ALGORITHMS

The operation of a GA is best explained by the use of a small and simple example. Consider a polluted aquifer that has two possible pumping locations, four possible pumping rates for each pump including not pumping, and two time intervals in the pumping regime allowing for one change of pumping rate. There are therefore 256 different possible pumping scenarios possible. Consider now how a possible solution can be encoded as a binary string. To encode four pumping rates requires two digits for each pump. For example:

0 0 = no pumping   0 1 = rate 1
1 0 = rate 2   1 1 = rate 3

For two pumping periods, with different pumping rates in each period, requires four digits per pump. For example, 0 1 1 1 can be separated into two groups of two digits: 0 1 and 1 1. The first pair represents the pumping rate in time period 1 and the second pair the rate in time period 2. In this example the pump is pumping at rate 1 during time period 1 and rate 3 in time period 2. With two pumps involved a total of eight digits are required. For example 0 1 1 1 0 0 1 0 can be separated into two groups of four digits: 0 1 1 1 and 0 0 1 0. The first group represents the first pump and the second group the second pump, which is not pumping during the first time period and at pumping rate 2 during the second time period.

The stages in the GA are then as follows.

2.1 Stage 1 (Selection)

A number of solutions are randomly generated. For this problem four solutions will be used and each is represented by a string of digits. Each of these solutions is considered a 'parent' in genetic terms and each of the digits a 'gene'. Assume the selected strings were:

01110010	parent 1
11101110	parent 2
00010111	parent 3
10101010	parent 4

2.2 Stage 2 (Analysis)

Each solution to the remediation, as defined by a parent, is analysed for efficiency defined by cost. For this paper finite element analysis will be used. However, any appropriate analysis may be used. If the resulting analyses generate the following costs:

parent 1 = \$10000   parent 2 = \$11000
parent 3 = \$9000;   parent 4 = \$12000

then obviously parent 3 is the best solution and parent 4 the worst.

2.3 Stage 3 (Tournament)

Pairs of parents are selected on a random basis for a competition to select the best solution from each pair. The numbers of pairs are equal to the number of parents and the selection procedure ensures each parent appears twice in the tournament. For example if the pairs were as shown below:

Competitor 1	Competitor 2	Winner
1	3	3
2	1	1
3	4	3
4	2	2

then the winner is shown in the right hand column. As can be seen, unless the best parent competes with itself it will appear twice in the winning list. Also, unless the worst parent competes with itself it will be eliminated.

2.4 Stage 4 (Mating)

Pairs of winners from the tournament are randomly selected for mating. A random number is generated and if it is less than a defined probability of mating then mating occurs, otherwise the parents proceed into the next generation unchanged. Using the winning parents 2 and 3 to illustrate the process a crossover location on the gene strings is randomly selected. If the selected location was after the fifth gene, then the gene strings are cut and spliced as shown below. The genes of one parent have been highlighted for clarity.

11101 110 = parent 2   **11101**111 = child 1
00010 111 = parent 3   00010**110** = child 2

This then generates two new solutions to proceed into the next generation each containing information from each of the parents. This hopefully will produce at least one solution that is more efficient than either of the parents. It may of course generate two less efficient solutions.

2.5 Stage 5 (Mutation)

Each gene of each of the children is considered in turn. A random number is called and if the number is less than a defined probability of mutation the gene digit is switched to bring variation into the optimisation process. This ensures that the process has a chance of not converging onto a local optimum as at least the mutated solution jumps into a different part of the search space. For example if the third gene of child 2 was mutated the result would be:

00010110 >> 00110110

The resulting children then form the parents of the next generation and the process is repeated until, over many generations, there is no improvement in efficiency.

3. THE OBJECTIVE FUNCTION

The objective function (Cost) for the GA to optimise consists of four parts:

$$\text{Cost} = a + b + c + d$$

where:
a is the cost of installing a well
b is the cost of each volume of polluted water extracted
c is a cost per unit level of pollutant above the permitted level at each node of the finite element mesh. This cost is to differentiate between different solutions that have exactly the same number of pumps and volumes of water extracted.
d is a penalty function that is applied if the pollutant level at any node is above the permitted level. This cost is to ensure that the GA is forced to reduce the pollutant level to below the accepted level, rather than selecting a solution that leaves high pollutant levels, but is cheaper than one that reduces the pollutant level to below the accepted level.

4. AN ILLUSTRATIVE EXAMPLE

Consider the two region confined aquifer shown in Figure 1. The aquifer was modelled with eight noded isoparametric elements and covered an area of 600m by 2000m The aquifer coefficients are identical for both regions except for the aquifer thickness. The coefficients used in the analysis were:

Thickness region 1	5m
Thickness region 2	10m
Storage	0
Longitudinal dispersivity	150m
Transverse dispersivity	100m
Conductivity	20m/d

The effect of elastic storage on the flow has been ignored for simplicity. However, it could just as easily have been included and transient conditions still exist due to pumping rate changes and through dispersion for the contaminant transport.

The hydraulic boundary conditions are a head difference of 5m between the left hand side and the right hand side of the aquifer. A zero pollutant concentration is specified on all boundaries. At time t=0 a pollutant is injected into the aquifer and after 50 days the pollutant has dispersed until the 10mg/l contour is as shown in Figure 1. This is the maximum acceptable level of pollutant. The object is to reduce the pollutant to below the maximum permitted level everywhere in the aquifer.

The eleven possible well positions along the line of symmetry are shown in Figure 1. The extraction period is limited to 100 days and has been divided into three sub-periods of 30, 30 and 40 days for pumping rate changes. Each pump can pump at 35, 70 or 105m^3/d or not be pumping. The objective function costs are:

• Pump installation	\$7000
• Pollution treatment per m^3	\$5
• Penalty per mg/l excess of the permitted level per node	\$100
• Global penalty for exceeding the permitted level anywhere	\$1000000

Using 200 parents with 50 generations and crossover and mutation coefficients of 0.9 and 0.01 respectively, the GA suggested the optimum solution shown in Table 1, which was generated after 46 generations. The cost of this solution was \$172000.

To verify that this is an optimum, the pumping rate at each of the suggested pumps was reduced in turn to the next lower level. In each case the pollutant level exceed the permitted level somewhere in the aquifer, indicating that the solution was at least a local optimum.

The suggested pumping regime is sensible as it introduced pumps further to the right as the pollutant plume moves to the right. At the same time it turns off pumps on the left hand side.

To illustrate the power of the technique, consider the same problem but with the pumping rates increased to 40, 80 and 120m^3/d. The problem set for the GA was to reduce the pollutant level to below the acceptable limit within 100 days. In this case the GA determined that it could be most economically done in 60 days with the solution shown in Table 2. Not only is the extraction time reduced but the cost is also reduced to \$132000, as less pumps are used and less water is removed for treatment, due to higher concentrations of pollutant being removed compared with the reduced pumping rates.

5. COMMENTS

The GA optimisation of the extraction process as discussed above has been applied to various theoretical aquifers, homogeneous and spatially varying. In each

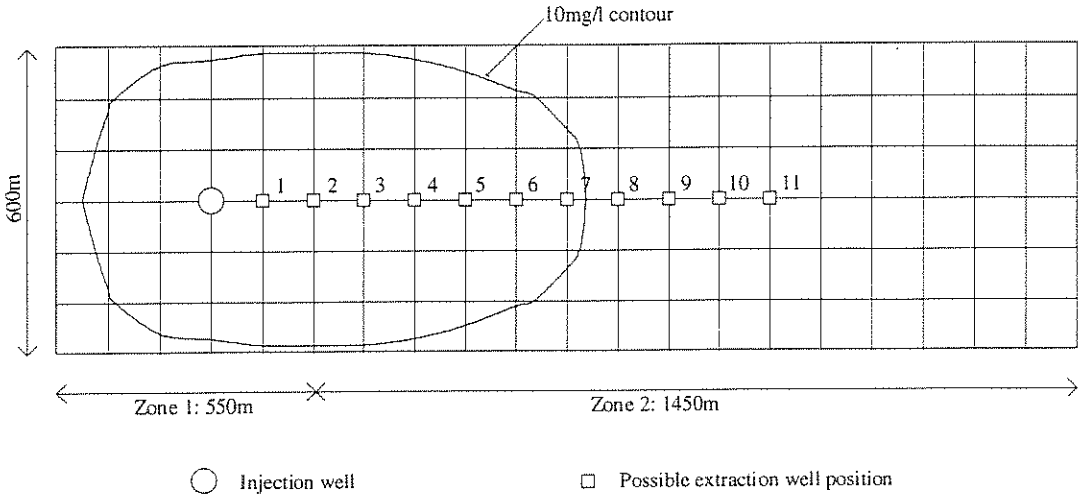

Figure 1: Two zone aquifer showing the location of contaminant injection possible extraction well locations and the 10mg/l contour prior to extraction commencing.

Table 1: Suggested optimum pumping regime to reduce the pollutant levels to the permitted level using pumping rates of 35, 70 and 105m^3/d.

Well Numbers	Time Period		
	1 (m^3/d)	2 (m^3/d)	3 (m^3/d)
1	105		
2	105		
4	105		
5		105	
6		105	
7			35
8		70	35
10		35	35

Table 2: Suggested optimum pumping regime to reduce the pollutant levels to the permitted level using pumping rates of 40, 80 and 120m^3/d.

Well Numbers	Time Period		
	1 (m^3/d)	2 (m^3/d)	3 (m^3/d)
1	120		
2	120		
6		80	
7		80	
8		80	
9		120	

case the GA determined a sensible solution that was realistic in so much as the pumping regime followed the pollutant plume. Occasionally a pump would be turned on early and was pumping relatively unpolluted water. Investigation of this situation determined that if this pump was turned off or the pumping rate was reduced the pollutant level was not reduced everywhere to the required level. The action of this pump was to increase the water velocity and bring the pollutant plume in range of other pumps that were active. This was in addition to increasing the dispersion due to the increased velocity.

When the time varying pumping regime was changed to a constant pumping regime it was found that the cost of remediation increased by up to 25%. Increasing the number of time intervals can increase or decrease the costs. The increase can be caused by the GA having to pump longer from a particular pump than is optimum. For example if the time periods for the illustrative problem were 20, 20, 30 and 30 days, then the pumps that were active for the initial 30 day period now have to be pumping for either 20 or 40 days. This could be detrimental to the efficiency.

The effect of introducing a two dimensional grid of possible extraction wells has also been investigated. With this arrangement it is possible to concentrate more wells nearer to the injection point. It is also possible to place a line of extraction wells across the path of the pollutant plume. Two dimension grids are usually more efficient than a one dimensional array with savings up to about 15%.

One interesting aspect of two dimensional grids is that if a symmetrical problem is posed the solution is not always symmetrical. Frequently a well will be turned on without the symmetrical well being activated. Investigations revealed that it was usually the case that the asymmetry was caused by the pumping rate selection. For example if the pumps in the illustrative example were used, then an asymmetric pump pumping at $35m^3/d$ may be turned on because this increase in volume is all that is required to generate an acceptable solution. The introduction of the symmetric pump would only have increased the cost without assisting the solution.

6. CONCLUSIONS

Evolutionary solution techniques when used to determine the extraction process of removing pollutant from an aquifer generate sensible engineering solutions. There is no guarantee that the generated solution is the global optimum solution but it will be of acceptable efficiency that can be implemented without further analysis. Alternatively the generated solution can be used as a starting point for further investigation by another technique.

7. REFERENCES

Cieniawski, S.E., Eheart, J.W. and Ranjithan, S. Using genetic algorithms to solve a multi-objective ground water monitoring problem, *Water Resources Research*, 31(2), pp 399-409, 1995.

Goldberg, D.E. *Genetic algorithms in search, optimisation and machine learning*, Addison-Wesley, 1989.

Holland, J.H. *Adaptation in natural and artificial systems*, University of Michigan Press, Ann Arbour , 1975.

McKinney, D.C. and Lin, M-D, Genetic algorithm solution of ground water management models. *Water Resources Research*, 30(6), pp 1897-1906, 1994.

Ritzel, B.J. and Eheart, J.W. Using genetic algorithms to solve a multi-objective ground water pollution containment problem. *Water Resources Research*, 30(5), pp 1589-1603, 1994.