# Aquifer Parameter Identification Using Evolutionary Processes

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

**Abstract** The accuracy of any groundwater analysis depends on the correct specification of the aquifer constants: the transmisivity, storage and porosity for the flow and the dispersivities for the contaminant movement. These coefficients are usually determined from simple pumping and tracer tests. For relatively uncomplicated aquifers there are analytical and graphical procedures that enable determination of the required coefficients. For aquifers involving spatial variations in the coefficients, numerical optimisation techniques are required. Gradient following techniques have limitations, however, recent developments in evolutionary techniques have overcome these limitations. A genetic algorithm approach produce sensible values of the coefficients but necessitate a series of analyses with the parameter limits being refined with each analysis until a satisfactory convergence is obtained. A recently reported technique known as the Shuffled Complex Evolution (SCE) incorporates four concepts: (a) the combination of probabilistic and deterministic approaches; (b) clustering; (c) systematic evolution of a complex of points spanning the parameter space, and (d) competitive evolution. This technique incorporates both the gradient following and genetic algorithm approaches and has been applied in this paper to the determination of the aquifer constants. It has proved to be more efficient than a simple genetic algorithm approach and is robust and effective. Both synthetic and simulated field data are used to illustrate the technique for anisotropic and spatially varying aquifers. The results of this study indicate that the SCE technique is efficient and can be used as an alternative to traditional gradient based methods normally used in groundwater parameter identification and is much more robust than genetic algorithms used alone.

## 1. INTRODUCTION

Field problems involving flow and solute transport in porous media require both forward (simulation) and backward (calibration) analyses for problem solution. The backward solution, otherwise known as the inverse problem, is required to determine the physical constants involved in the problem by attempting to fit observed data changes caused through known system perturbations. The forward solution then uses these determined constants to predict the changes to the system through changing the boundary conditions including injection or extraction.

The forward solution presents very few problems as changes to an existing system's conditions, caused through system perturbations can be readily determined through finite element or finite difference analyses. This applies irrespective of the complexity of the system provided it can be accurately defined. The inverse problem, however, presents significant difficulties in all but the most trivial situations. These difficulties are caused by having to use an optimisation technique to determine the values of the constants that provide a best match of the calculated system changes to the measured changes. Sun, [1994] listed the difficulties as: (1) the problem may be non-unique and unstable with respect to observation errors, (2) the quantity and quality of the observed data is usually inadequate, and (3) the error in modelling can dominate all other errors. Duan et al [1992] also defined five difficulties encountered by an optimisation technique utilised in inverse problems: (1) there may be several major regions of attraction into which a search strategy may converge; (2) each major region of attraction may contain numerous local minima or maxima which occur both close to and at various distances from the best solution; (3) the objective function surface in the multi-parameter space may not be smooth and may not even be continuous, and the derivatives may be discontinuous and may vary in an unpredictable manner through the parameter space; (4) the parameters may exhibit varying degrees of sensitivity and a great deal of interaction and compensation, and much of the interaction can be highly nonlinear; and, (5) the response surface near the true solution is often nonlinear. These difficulties, in determining the parameters involved in calibrating watershed models, have been overcome by Duan et al [1992,1993] through the use of a new optimisation technique called the shuffled complex evolution technique (SCE). This paper applies the SCE to groundwater problems.

## 2. THE SHUFFLED COMPLEX EVOLUTION TECHNIQUE

The SCE technique is based on four concepts: (1) combination of deterministic and probabilistic approaches, (2) systematic evolution of a complex of points spanning the parameter space in the direction of global improvement, (3) competitive evolution and (4) complex shuffling.

The synthesis of these four elements makes the SCE method not only effective and robust but also flexible and efficient. The use of deterministic strategies permit the SCE to make effective use of response surface information to guide the search. Robustness and flexibility are taken care of through the inclusion of random elements. The process of concentrating a search in the most promising region of the search space is guided

by an implicit clustering strategy. The use of a systematic complex evolution strategy helps to ensure a relatively robust search that is guided by the structure of the object function.

The SCE technique works in the following way:

- A number (s) of possible solutions is generated randomly. The values of the parameter coefficients are randomly selected between realistically specified limits.
- Each of the 's' solutions is ranked according to some criterion to measure the accuracy of the solution.
- The 's' solutions are partitioned into 'p' sub-groups or complexes each containing 'm' solutions such that complex number 'i' contains solutions (p(k-1)+i) from the ranked list, where k=1,m and i=1,p.
- Each complex is evolved according to the competitive complex evolution (CCE) algorithm described below.
- The complexes are recombined to generate a single ranked list again and then the 'p' complexes recreated but now with new members.
- Convergence is checked and if not satisfied the CCE is repeated.

The stages in the competitive complex evolution algorithm for each complex are:

(1) In each complex randomly create a sub-complex of 'q' solutions according to a trapezoidal probability distribution such that the best solutions have the highest chance of selection.
(2) Identify the worst solution of the 'q' solutions and determine the centroid of the parameter values of all the other (q-1) points.
(3) Generate a reflection solution by reflecting the parameters of the worst point through the centroid. If the reflection solution is in feasible space go to stage 4. Otherwise go to stage 6
(4) If the newly generated solution is better than the worst solution, replace the worst solution with the new solution and go to stage 7. Otherwise go to stage 5
(5) Generate a contraction solution by determining the parameters mid way between the centroid and the worst solution. If the contraction solution is better than the worst solution, replace the worst solution with the contraction solution and go to stage 7. Otherwise go to stage 6.
(6) Randomly generate a new solution and replace the worst solution.
(7) Repeat stages 2 to 6 'a' times where 'a' is the number of new solutions to be generated for each sub-complex.
(8) Repeat stages 1 to 7 'b' times where 'b' is the number of evolution steps taken by each complex before the complexes are recombined and shuffled.

The values for m, p, q, a and b are user specified. However, values of : m=2n+1, q=n+1, a=1 and b=m produced sensible solutions for most problems, where n is the number of parameters to be determined. The value of p depends on the degree of difficulty of the problem. The more difficult the problem the more complexes needed to locate the global optimum.

## 3. THE FLOW COEFFICIENTS

The accuracy of the SCE in determining the flow coefficients was tested using the two dimensional aquifer model shown in Figure 1, which covered a 1000m by 1000m area. The aquifer was modelled using finite element analysis and the mesh consisted of 100 four noded elements and 121 nodes. Spatial variation was imposed by defining six different regions in the aquifer with three different aquifer zones, each with its own set of aquifer properties. The numbers in each of the elements define which zone is associated with each region. A 50m head was specified on all boundaries and a single extraction well and four test wells were utilised for the determination of the coefficients for each of the three aquifer zones.

Each of the three aquifer zones was assumed to be anisotropic. The aquifer coefficients used to generate the synthetic drawdown data through a forward analysis are shown in Table 1. The data supplied to the SCE was: (1) the drawdown data at the test wells at 10 specified time intervals; (2) the reasonable limits for the flow coefficients, shown in Table 1; (3) the extraction pumping rate; and (4) the model of the aquifer as defined in Figure 1. To simulate field data the synthetic data was randomly corrupted by up to 100mm for each head reading supplied to the SCE. The head drop at the test wells was up to 4.384m over a time period of 10 days. For this problem 10 complexes were used with 25 solutions per complex giving 250 initial trial solutions. The number of cycles required to complete the analysis was 182 which resulted in a total of 62,581 function evaluations.

The results of the SCE analysis are shown in Table 1. For the synthetic data the aquifer flow coefficients were determined to a high order of accuracy, with an average drawdown head error of $2.929 \times 10^{-10}$ m per reading For the simulated field data the Kxx, Kyy and the storage coefficients (S) were accurately determined. However, the accuracy of Kxy was not quite as good but the average head error was still only $1.060 \times 10^{-5}$ m. This reduction in accuracy was almost certainly due to the magnitude of Kxy being only 5% of the value of Kxx and Kyy but mainly due to the corruption of the synthetic data to represent field errors. The solution is therefore highly sensitive to minor errors in the head readings and

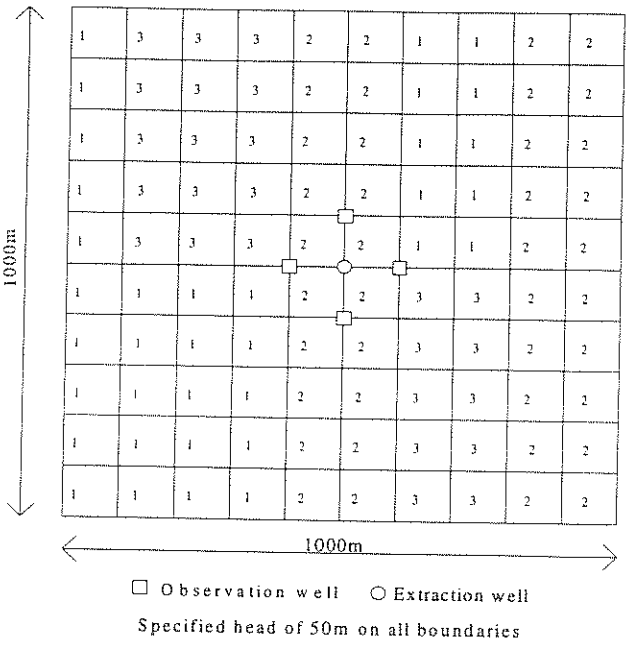


**Figure 1:** Finite element mesh for flow coefficient determination showing aquifer spatial variation

**Table 1:** The theoretical values, limits and the values of the conductivities and storage coefficient determined by the SCE technique

| Zone | Kxx | Kyy | Kxy | S |
|---|---|---|---|---|
| Limits for all zones | 0.0-150 | 0.0-150 | 0.0-150 | 0.0-1.0 |
| 1: theoretical | 40 | 30 | 1.5 | 0.002 |
| : synthetic data | 40.075 | 29.993 | 1.5777 | 0.002004 |
| : field data | 43.408 | 30.684 | 4.9974 | 0.00211 |
| 2: theoretical | 30 | 20 | 1.0 | 0.003 |
| : synthetic data | 30.0 | 20.0 | 0.9916 | 0.003 |
| : field data | 30.022 | 19.978 | 2.2148 | 0.003008 |
| 3: theoretical | 50 | 35 | 1.75 | 0.001 |
| : synthetic data | 49.927 | 35.994 | 1.8454 | 0.000998 |
| : field data | 49.708 | 34.665 | 3.45 | 0.000992 |

the relative magnitudes of the conductivities. But, with the small differences in the heads between the simulated field data and those calculated using the coefficients as determined by the SCE, any Hydrologist would accept that the SCE had determined the correct coefficients.

It is unlikely that the head error could have been decreased by increasing the numbers of trial solutions or complexes due to the corruption of the synthetic data to generate the simulated field data.

During the optimisation if, after a reasonable number of cycles, the head error had not reduced to a sufficiently low enough value for convergence to have occurred, then it is possible that the limits of the coefficients in which the SCE had to work were inappropriate. For example if Kxy had been negative for some of the aquifer zones, then it would not have been possible for the SCE to have obtained convergence with the limits set. In this situation the best match determined by the SCE would almost certainly have included one or more of the coefficients equal to one of the limits that had been set, or at least very close to that limit. This would have been a strong indication that the limits were inappropriate.

Yeh [1986] indicated that instability in gradient based optimisation techniques, used for inverse groundwater problems, could be avoided if the number of test wells were greater than the number of parameters to be determined. In this problem 12 parameters were determined using 4 test wells. The SCE therefore appears to be superior to gradient based techniques as it does not require the calculation of derivatives during the search procedure.

The SCE has been applied to various types of confined aquifer. For a homogeneous and anisotropic aquifer the correct coefficients were detected with only 2 complexes for both the synthetic and simulated field data with a maximum head error of $1.67x10^{-8}$. For a three dimensional aquifer consisting of three layers the maximum average head error was $1.118x10^{-5}$ for the simulated field data. Other multi-zoned anisotropic aquifers have also been analysed with similar results.

Similar problems to that described above, in both two and three dimensions, have been analysed and have confirmed that the SCE is stable and robust for these problems.

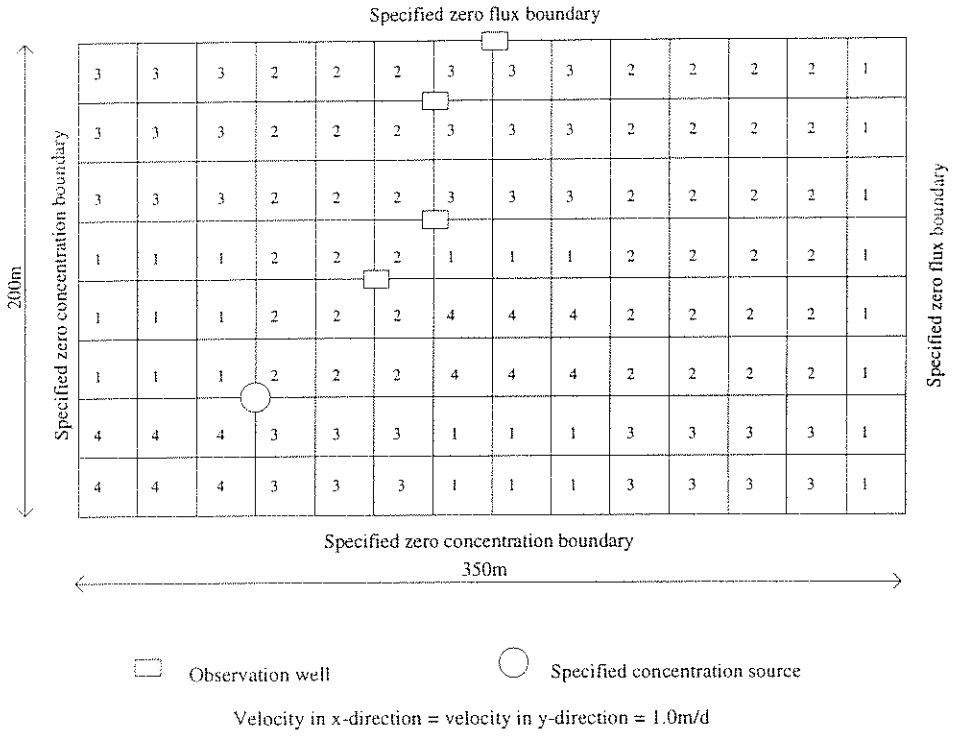


**Figure 2:** Finite element mesh for contaminant transport coefficients determination showing aquifer spatial variation

**Table 2:** Theoretical values, limits and the values of the contaminant transport coefficients determined by the SCE technique

| Zone | Longitudinal dispersivity | Transverse dispersivity | Porosity |
|---|---|---|---|
| limits for all zones | 0.0-250 | 0.0-250 | 0.0-1.0 |
| 1: theoretical value | 150 | 120 | 0.25 |
| : synthetic data | 150.01 | 120.0 | 0.25001 |
| : field data | 150.04 | 119.81 | 0.24978 |
| 2: theoretical values | 135 | 105 | 0.15 |
| : synthetic data | 134.99 | 104.99 | 0.14999 |
| : field data | 136.80 | 106.23 | 0.15195 |
| 3: theoretical values | 125 | 100 | 0.35 |
| : synthetic data | 125.00 | 99.998 | 0.34999 |
| : field data | 124.42 | 99.748 | 0.34767 |
| 4: theoretical values | 145 | 115 | 0.45 |
| : synthetic data | 145.09 | 115.03 | 0.45016 |
| : field data | 135.15 | 110.72 | 0.43046 |

## 4. THE CONTAMINANT TRANSPORT COEFFICIENTS

A similar problem to that used in Section 3 above was used to illustrate the application of the SCE technique to determining the contaminant transport coefficients. The finite element mesh is shown in Figure 2 and consisted of

112 four noded elements and 135 nodes and covered an area of 200m by 350m. The average fluid velocity in each of the coordinate directions was 1m/day and the area was divided into 12 zones with 4 different aquifer zones as shown. The contaminant source was a specified concentration of 100mg/l at a defined location and all the boundary conditions including this point source are also shown in Figure 2.

With the complexity of this problem the search space was divided into 10 complexes with 25 solutions per complex giving a total of 250 initial trial solutions. The number of cycles required to complete the analysis was 131 which resulted in a total of 45979 function evaluations.

The true parameters and the specified limits in which the SCE was to work are shown in Table 2. The contamination levels at the four observation wells were specified at 10 time intervals over a period of 50 days. The contaminant levels for the synthetic data reached a maximum level of 22.7mg/l during this time period. To simulate field data errors the synthetic data was randomly corrupted by up to 0.2mg/l for each of the readings supplied to the SCE.

The coefficients selected by the SCE are shown in Table 2. As in the previous example the agreement with the synthetic data is excellent with an average error of $2.47 \times 10^{-10}$ mg/l per reading. For the field data the average error was $6.66 \times 10^{-6}$ mg/l per reading. The simulated field data coefficients had small variations compared with the theoretical coefficients for the same reasons as discussed in Section 3.

For similar reasons as those given for the flow coefficients, it is unlikely that the contaminant level error could have been decreased by increasing the numbers of trial solutions or complexes due to the corruption of the synthetic data to generate the simulated field data.

The technique has been applied to other aquifers both single zoned and multi-zoned and two and three dimensional. The average contaminant reading error for each problem was of a similar magnitude to those above.

## 5. CONCLUSIONS

The SCE technique has proved that it can determine the aquifer coefficients to acceptable levels of accuracy when simulated field data was used. It must therefore now be considered as a possible alternative to other methods such as gradient based techniques that are currently being used. It does not exhibit the instability of gradient based techniques as it does not use the derivatives of the function to be minimised but uses the function directly.

The significant computer time involved with the technique could be reduced through the use of parallel computers so that the evolution of each of the complexes occurs concurrently rather than sequentially.

## 6. REFERENCES


Duan, Q., S. Sorooshian and V.K. Gupta, Effective and efficient global optimisation for conceptual rainfall-runoff models, *Water Resour. Res.*, 28(4),1015-1031, 1992

Duan, Q., V.K. Gupta S and S. Sorooshian, A shuffled complex evolution approach for effective and efficient global minimisation, *J. Optim. Theory Appl.*, 76(3), 501-521, 1993

Sun, N.Z., Inverse problems in groundwater modelling, Kluwer Academic Press, Boston, 1994

Yeh, W.W-G., Review of parameter identification procedures in groundwater hydrology; the inverse problem, *Water Resour. Res.*, 22(2), 95-108, 1986