Simultaneous estimation of Aquifer Parameters and Parameter Zonations using Genetic Algorithm

M.Tamer AVVAZ
Visiting Graduate Student

Nov 20/2006
MULTIMEDIA ENVIRONMENTAL SIMULATIONS LABORATORY (MESL)
School of Civil and Environmental Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332, USA
OUTLINE

Objective
Model Development
  Parameter Estimation and Inverse Models
  Simulation
  Optimization with Genetic Algorithms
  Pattern Classification
  Problem Formulation and Search Procedure
Numerical Example
Conclusions
Objective
Model Development
  Parameter Estimation and Inverse Models
  Simulation
  Optimization with Genetic Algorithms
Pattern Classification
  Problem Formulation and Search Procedure
Numerical Example
Conclusions
Objective

- Mathematical simulation models are used to analyze and determine the sustainable usage and management of groundwater systems.
- These models require many hydrogeologic parameters including transmissivity, storativity, specific yield and recharge/infiltration.
- These aquifer parameters are generally obtained from laboratory experiments and/or field studies.
- Measurement and estimation of these spatially distributed parameters in large aquifer systems are time consuming and costly.
- In general, obtaining the hydraulic heads field data is relatively easier and cheaper than obtaining other parameters.
- Estimation of aquifer parameters based on limited field observations (for example piezometric head) is known as inverse modeling.
Objective

Aquifer Properties:
Hydrologic, Hydrogeologic, Water Quality, etc.

Water Demand:
Drinking, Irrigational, Industrial, etc.

Solution of Partial Differential Equations
Initial-Boundary Conditions, Sinks/Sources

Numerical Solution:
Finite Diff., Finite Elem., Boundary Elem., etc.

Model Verification

Aquifer's Response for Different Management Scenarios

Evaluation for Quality and Quantity (Sustainable Usage)

must be known
Objective

Model Development
  Parameter Estimation and Inverse Models
  Simulation
  Optimization with Genetic Algorithms
  Pattern Classification
  Problem Formulation and Search Procedure

Numerical Example

Conclusions
Model Development

Parameter Estimation and Inverse Models

- In order to solve groundwater problems, two methods are widely used:
  - Forward Model
  - Inverse Model

- The forward models are used to obtain hydraulic heads for given aquifer parameters (i.e. transmissivity, storativity, etc.)
- The inverse models are used to back out underlying model parameters using the hydraulic heads.
**Parameter Estimation and Inverse Models**

- There are several inherent difficulties with inverse solution techniques.
- One of the major difficulties in inverse modeling problems is their ill-posed nature.
- Ill-posed nature is usually characterized by the instability in inverse numerical solutions and non-uniqueness of the identified parameters.
- The instability and non-uniqueness of the inverse problems occur when small errors are introduced to the observed values (or numerical errors).
- This situation causes large errors in the values of the parameters identified.
Model Development

Parameter Estimation and Inverse Models

- The relation between hydraulic heads and transmissivities for both forward and inverse models:

**Forward Model**

\[ H = f(T) + \varepsilon \]

\[ \overline{T} = g(H) \]

**Parameter Space**

**Inverse Model**

\[ f : \text{Forward Operator} \]

\[ \varepsilon : \text{Error Vector} \]

\[ g_i : \text{Generalized Inverse Operator} \]

\[ \varepsilon \to 0 \Rightarrow T = \overline{T} \]
Parameter Estimation and Inverse Models

Several solution techniques have been used in order to solve inverse problems during the past two decades

The first group of solutions:
Statistical solution methods and Recursive filtering techniques

The second group of solutions:
Combination of several optimization approaches with numerical solution techniques
OUTLINE

Objective

Model Development
  Parameter Estimation and Inverse Models
  Simulation
    Optimization with Genetic Algorithms
    Pattern Classification
    Problem Formulation and Search Procedure

Numerical Example

Conclusions
Simulation

Simulation phase can be defined as the numerical solution of the governing partial differential equation for a given initial and boundary conditions.

The governing equation that describes flow in a two-dimensional, unconfined, compressible, non-homogeneous aquifer may be given as:

\[
S \frac{\partial h}{\partial t} = \frac{\partial}{\partial x} \left( T_{xx} \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T_{yy} \frac{\partial h}{\partial y} \right) \pm W
\]
Model Development

**Simulation**

- The following finite difference equation for heterogeneous material properties and non-uniform grid scheme is used for the solution:

\[
\frac{h_{i,j}^{n+1}}{h_{i,j}^{n+1}} = \left[ \frac{CC.(h_{i,j}^{n+1}) + CE.(h_{i+1,j}^{n+1}) + CW.(h_{i-1,j}^{n+1}) + CS.(h_{i,j+1}^{n+1}) + CN.(h_{i,j-1}^{n+1}) + W_{i,j}}{(CE + CW + CS + CN + CC)} \right]
\]

- \( CE = \frac{2T_{i,j}T_{i-1,j}}{(\Delta x_{(i)}T_{i+1,j} + \Delta x_{(i+1)}T_{i,j})\Delta x_{(i)}} \)

- \( CW = \frac{2T_{i,j}T_{i-1,j}}{(\Delta x_{(i-1)}T_{i+1,j} + \Delta x_{(i+1)}T_{i,j})\Delta x_{(i)}} \)

- \( W_{i,j} = \pm \frac{Q_{i,j+1}^{m+1}}{\Delta x_{(i)}\Delta y_{(j)}H_{i,j}} \)

- \( CC = \frac{S_{i,j}}{\Delta t} \)

- \( CN = \frac{2T_{i,j}T_{i,j-1}}{(\Delta y_{(j-1)}T_{i,j} + \Delta y_{(j)}T_{i,j-1})\Delta y_{(j)}} \)

- \( CS = \frac{2T_{i,j}T_{i,j+1}}{(\Delta y_{(j)}T_{i+1,j} + \Delta y_{(j+1)}T_{i,j})\Delta y_{(j)}} \)
OUTLINE

Objective
Model Development
  Parameter Estimation and Inverse Models
  Simulation
  Optimization with Genetic Algorithms
Pattern Classification
Problem Formulation and Search Procedure
Numerical Example
Conclusions
Model Development

*Optimization with Genetic Algorithms*

Genetic Algorithms:
- efficient in continuous and discrete optimization
- gives good results in finding global or near global optimum solutions
- does not require to evaluate the derivatives of the fitness function
- does not require the information on an initial solution
- finds the solution as a random search based on “survival of the fittest” concepts of the GA.

In GA, variables are usually represented by binary string chromosomes

Each chromosome is evaluated by using genetic operators (Direct Selection, Selection, Crossover and Mutation)
Model Development

**Optimization with Genetic Algorithms**

<table>
<thead>
<tr>
<th>Variable (Decimal)</th>
<th>Variable (Binary String)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>$X_1$</td>
<td>1010...01</td>
</tr>
<tr>
<td>$X_2$</td>
<td>110...10</td>
</tr>
<tr>
<td>$X_3$</td>
<td>010...01</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Model Development

**Optimization with Genetic Algorithms**

**Selection (Roulette Wheel)**

*Before Selection*

<table>
<thead>
<tr>
<th>1</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

*After Selection*

<table>
<thead>
<tr>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Model Development

Optimization with Genetic Algorithms

Crossover (Two-Point)

Chromosome A: 1 1 1 0 0 1 0
Chromosome B: 1 0 0 1 0 0

Chromosome A’: 1 1
Chromosome B’: 1 0
Model Development

*Optimization with Genetic Algorithms*

*Mutation (Adaptive)*

| Chromosome A' | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
Model Development

*Optimization with Genetic Algorithms*

*Diversity Control of Population*

While Number of Generation > Stall Generation Limit

If \( \frac{f_j - f_{j-st}}{f_j} \leq \varepsilon \) Then

Add initially generated chromosomes to the population by saving the elite values

End if

Wend

\( f_j \) : Fitness at the \( j \)th generation

\( f_{j-st} \) : Fitness at the \((j-st)\)th generation

\( st \) : Stall generation limit

\( \varepsilon \) : Tolerance
Model Development

Optimization with Genetic Algorithms

Error Indicators

Modified Coefficient of Efficiency ($E$)

$$
E = 1 - \frac{\sum_{l=1}^{L} |h_{\text{obs},l} - h_{\text{est},l}|}{\sum_{l=1}^{L} |h_{\text{obs},l} - \bar{h}_{\text{obs}}|}
$$

- $h_{\text{obs}}$: Observed hydraulic heads
- $h_{\text{est}}$: Estimated hydraulic heads
- $\bar{h}_{\text{obs}}$: Mean of the observed hydraulic heads
- $L$: Number of observation wells
OUTLINE

Objective

Model Development

  Parameter Estimation and Inverse Models
  Simulation
  Optimization with Genetic Algorithms

Pattern Classification

  Problem Formulation and Search Procedure

Numerical Example

Conclusions
Model Development

**Pattern Classification**
- Patterns can be classified based on the values of membership functions.

- For example, in a one-dimensional problem with two pattern types, two density functions of normal distribution can be used to separate the two groups.

\[
N_i(X, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ -\frac{1}{2} \left( \frac{X - \mu_i}{\sigma_i} \right)^2 \right] \quad i = I \& II
\]

- $\mu_i$ : Mean and the location of the centroid of pattern $i$
- $\sigma_i$ : Standard deviation
- $X$ : Position of the grid point
Model Development

Pattern Classification

Two-dimensional problem can use two dimensional density function of normal distribution to determine patterns as follows:

\[
N_i(X, Y, \mu_{Xi}, \mu_{Yi}, \sigma_{Xi}, \sigma_{Yi}) = \frac{1}{2\pi \sigma_{Xi}\sigma_{Yi}} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{X - \mu_{Xi}}{\sigma_{Xi}} \right)^2 + \left( \frac{Y - \mu_{Yi}}{\sigma_{Yi}} \right)^2 \right] \right\}
\]

How can we determine the pattern???
Pattern Classification

A grid point \( P(X, Y) \) can be classified to pattern \( k \), if the \( N_k \) yields the maximum value among \( n \) functions as seen in the following:

\[
P(X, Y) \in \text{Pattern } k, \text{ if } N_k = \text{Max}[N_1, N_{II}, \cdots N_k, \cdots, N_n]
\]

Variables of the each pattern to be optimized: \((\mu_{Xi}, \mu_{Yi}, \sigma_{Xi}, \sigma_{Yi})\)

How does it work???
Model Development

Pattern Classification

\[ N_1 = f(\mu_{x1}, \mu_{y1}, \sigma_{x1}, \sigma_{y1}) \]

\[ N_2 = f(\mu_{x2}, \mu_{y2}, \sigma_{x2}, \sigma_{y2}) \]

\[ N_3 = f(\mu_{x3}, \mu_{y3}, \sigma_{x3}, \sigma_{y3}) \]
OUTLINE

Objective

Model Development
- Parameter Estimation and Inverse Models
- Simulation
- Optimization with Genetic Algorithms
- Pattern Classification
  Problem Formulation and Search Procedure

Numerical Example
Conclusions
Model Development

Problem Formulation and Search Procedure

The mathematical model can be summarized as follows:

\[
\text{Min } Z = \frac{1}{L} \sum_{t=1}^{N_t} \sum_{l} \left| h_{\text{obs},t,l} - h_{\text{est},t,l} \right|
\]

subject to

\[
S \frac{\partial h}{\partial t} = \frac{\partial}{\partial x} \left( T \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T \frac{\partial h}{\partial y} \right) \pm W \quad \bar{T} \in \{ T_1, T_2, \ldots, T_k, \ldots, T_N \}
\]

Decision Variables for Each Zone:

\[
\bar{T}_j = T_k \quad \text{if} \quad N_k = \text{Max}\{N_1, N_2, \ldots, N_k, \ldots, N_N\}
\]

\[
\left( T_i, \mu_{X_i}, \mu_{Y_i}, \sigma_{X_i}, \sigma_{Y_i} \right)
\]

\[
N(X,Y,\mu_{X_i},\mu_{Y_i},\sigma_{X_i},\sigma_{Y_i}) = \frac{1}{2\pi\sigma_{X_i}\sigma_{Y_i}} \exp \left\{ -\frac{1}{2} \left[ \left( \frac{X - \mu_{X_i}}{\sigma_{X_i}} \right)^2 + \left( \frac{Y - \mu_{Y_i}}{\sigma_{Y_i}} \right)^2 \right] \right\}
\]
Problem Formulation and Search Procedure

1. Start
2. Initial Solution
3. Determine pattern distribution and assign $T$ values to the patterns
4. Numerical Solution
5. Calculate Fitness Function
   - Fitness = $\min$
6. Selection (Roulette Wheel)
7. Crossover (Multi-point)
8. Mutation (Adaptive)
9. Direct Selection (Elitism)
10. Y: Go to End
11. N: Go back to Step 6
OUTLINE

Objective
Model Development
  Parameter Estimation and Invers Models
  Simulation
  Optimization with Genetic Algorithms
  Pattern Classification
  Problem Formulation and Search Procedure
**Numerical Example**
Conclusions
Numerical Example

Problem Definition

Solution Parameters

- Dimensions of Solution Domain: 6 km x 6 km
- Grid Spacing: 250 m
- Transmissivities: $T_1 = 150 \text{ m}^2/\text{day}$, $T_2 = 50 \text{ m}^2/\text{day}$, $T_3 = 200 \text{ m}^2/\text{day}$
- Recharge (Dashed Region): $0.00015 \text{ m/day}$
- Pumping Rates: $Q_1 = 800 \text{ m}^3/\text{day}$, $Q_2 = 3000 \text{ m}^3/\text{day}$, $Q_3 = 6000 \text{ m}^3/\text{day}$
- Storage Coefficient: $S = 0.01$
- Number of Observation Wells: 18
- Bit Number of Each Variable: 17
- Population Number ($p_z$): 100
- Direct Selection Rate ($p_d$): 0.10
- Mutation Rate ($p_m$): 0.01
- Crossover Rate ($p_c$): 0.85
- Maximum Number of Generation: 500
- Stall Generation Limit: 10
- Tolerance: 0.00001
Numerical Example

Three cases have been taken into account to test the performance of the proposed model

• Case 1 (Pattern known – Parameters unknown)
• Case 2 (Pattern unknown – Parameters known)
• Case 3 (Pattern unknown – Parameters unknown)
Numerical Example

Case 1 (Pattern known – Parameters unknown)
Numerical Example

Case 2 *(Pattern unknown – Parameters known)*
Numerical Example

Case 3 (Pattern unknown – Parameters unknown)

Initially generated population added by saving the elite chromosomes
OUTLINE

Objective
Model Development
  Parameter Estimation and Invers Models
  Simulation
  Optimization with Genetic Algorithms
  Pattern Classification
  Problem Formulation and Search Procedure
Numerical Example
Conclusions
Conclusions

Following conclusions can be drawn from this study:

• The proposed solution algorithm is applied to a hypothetical aquifer model to determine the transmissivity values and their patterns.

• While determination of aquifer parameters (inverse problem) is relatively easy, on the other hand determination of parameter zonations is more difficult.

• The value of the objective function is improved quickly in early iterations and then changes very slowly in subsequent iterations.

• Many neighbor solutions, which have similar values of decision variables, result in the same zonation and thus cause slow improvement of the value of objective function.
Conclusions

Following conclusions can be drawn from this study:

• One of the major difficulties to find the global optimum solution is the loosening of the diversity of the population pool.

• To satisfy the diversity of the population, initial population terms have been added to the pool by checking the stall generation limit.

• Results showed that after satisfying the diversity of population, variation of the fitness function increases.

• Model may be trapped to local optimum solutions when the number of population is small.

• After building the generalized solution technique, future study should be done on a real aquifer.
Contact Information:

**Address**: Department of Civil Engineering
Pamukkale University
Central Kampus
TR-20070 Denizli / TURKEY

**Phone**: (+90-258) 213-4030 / 1537

**Fax**: (+90-258) 212-5548

**E-mail**: tayvaz@pamukkale.edu.tr

**Web**: http://tayvaz.pamukkale.edu.tr