

Map Asia 2009, 18-20 August, 2009, Singapore  
“An integrated approach to Biogeography Based Optimization with  
case based reasoning for retrieving Groundwater possibility”  
Authors V. K .Panchal, Harish Kundra, Amanpreet Kaur  
Seminar Theme : **Technology Integration**

# An Integrated Approach to Biogeography Based Optimization with case based reasoning for retrieving Groundwater Possibility

<sup>1</sup> Harish Kundra, <sup>2</sup> Amanpreet Kaur, <sup>3</sup> V.K.Panchal

<sup>1</sup> Assistant Professor, RIEIT, Railmajra, Punjab; [kundra\\_harish@yahoo.com](mailto:kundra_harish@yahoo.com)

<sup>2</sup> Student, M-Tech, RIEIT, Railmajra, Punjab; [amanrandhawa85@gmail.com](mailto:amanrandhawa85@gmail.com)

<sup>3</sup> Scientist 'F', Defense Terrain Research laboratory, DRDO, Delhi ; [vkpans@ieee.org](mailto:vkpans@ieee.org)

## ABSTRACT

In this paper, we have proposed the implementation of Groundwater Possibility Retrieval System to detect the groundwater possibility using the burgeoning technique of Biogeography Based Optimization and Case based reasoning retrieval. In the case base used in this study, Geology, Land form, Soil, Land use, Slope and Lineament play a decisive role in determining ground water possibilities. These attributes are integrated and analyzed by the experts to predict the groundwater possibility. A stored case in case base is a preselected combination of these attributes having solution class. The cases reveal three solutions: high groundwater possibility, moderate possibility and low possibility. It is assumed that initially there are three habitats: High, Moderate and Low habitat. This is because according to BBO, solutions are analogous to the habitats. The cases are assumed to be the species that are residing in these habitats and their attributes are assumed to be the features of species. High, moderate and low habitats consist of species revealing high, moderate and low groundwater possibility respectively. Assume an input case (case without solution class) as an input species that is to be placed into an isolated habitat. Based on HSI, emigration takes place from other habitats to an isolated habitat. Isolated habitat is now consisted of species that are more similar to an input species. Case retrieval method is applied to find out the best match from the emigrated species to an input species. Thus solution class of a best matched species is transferred to an input species as its solution class.

**Keywords** – Biogeography, Biogeography Based Optimization, Case based Reasoning, Case retrieval, groundwater possibility.

## 1. INTRODUCTION

Biogeography is the study of the geographical distribution of biological organisms. The mindset of the engineer is that we can learn from nature. This motivates the application of biogeography to optimization problems. In this paper the mathematics of biogeography as the basis for the development of a new field: biogeography-based optimization (BBO) has been considered. We have proposed the concept of integration of BBO with case based reasoning retrieval to develop an information retrieval only system. Using BBO integrated with case retrieval method; we will retrieve the information regarding groundwater possibility in a given area.

### 1.1 Biogeography Based Optimization

Biogeography is the study of geographic distribution of biological organisms. The science of Biogeography can be traced to the work of century nationalists such as Algreed Wallace[1] and Charles Darwin[2]. BBO is an application of biogeography to optimization problems. It is modeled after the immigration and emigration of species between the islands. It is analogous to general problem solutions. Geographical areas that are well suited as residences for biological species are said to have a high habitat suitability index (HSI) [4]. Features that correlate with HSI include such factors as rainfall, diversity of vegetation, diversity of topographic features, land area, and temperature. The variables that characterize habitability are called suitability index variables (SIVs). SIVs can be considered the independent variables of the habitat, and HSI can be considered the dependent variable. Habitats with a high HSI tend to have a large number of species, while those with a low HSI have a small number of species. Habitats with a high HSI have many species that emigrate to nearby habitats, simply by virtue of the large number of species that they host. Habitats with a high HSI have a low immigration rate because they are already saturated with species, high HSI habitats are more static in their species distribution than low HSI habitats. Therefore By the same token, high HSI habitats have a high emigration rate; the large to emigrate to neighboring habitats number of species on high HSI islands has many opportunities. A good solution is analogous to an island with a high HSI, and a poor solution represents an island with a low HSI. High HSI solutions resist change more than low HSI solutions. By the same token, high HSI solutions tend to share their features with low HSI solutions. Poor solutions accept a lot of new features from good solutions. This addition of new features to low HSI solutions may raise the quality of those solutions. This new approach is called as problem solving BBO [3].

### 1.2 Case Based Reasoning

CBR involves reasoning from prior examples: retaining a memory of previous problems and their solutions and solving new problems by reference to that knowledge. Generally, a case-based reasoner will be presented with a problem, either by a user or by a program or system. The case-based reasoner then searches its memory of past cases (called the case base) and attempts to find a case that has the same problem specification as the case under analysis. If the reasoner cannot find an identical case in its case base, it will attempt to find a case or multiple cases that most closely match the current[5].

Based on an extensive survey of CBR[6], Aamodt and Plaza describe CBR as a four step process:

- *Retrieve* the most similar case or cases.
- *Reuse* the information and knowledge in that case to solve the problem.
- *Revise* the proposed solution if necessary.
- *Retain* the parts of this experience likely to be useful for future problem solving.

### 1.2.1 CBR Problem Areas

The challenge in CBR as elsewhere is to come up with methods that are suited for problem solving and learning in particular subject domains and for particular application environments. Core problems addressed by CBR research can be grouped into five areas. A set of coherent solutions to these problems constitutes a CBR method [6]:

- Knowledge representation
- Retrieval methods
- Reuse methods
- Revise methods
- Retain methods

Section 2 discusses an integrated approach to Biogeography Based Optimization with case based reasoning retrieval. Section 3 reviews the groundwater possibility detection and provides the results. Section 4 presents some concluding remarks.

## 2. AN INTEGRATION OF BBO WITH CASE BASE REASONING

In our research, we demonstrate the integration of BBO technique with Case Based reasoning Retrieval. The motivation of this research is to merge the burgeoning technique of BBO with Case retrieval method of Case Based reasoning in order to see how the two disciplines can be of mutual benefit. We have proposed the implementation of Groundwater Possibility Retrieval System to detect the groundwater possibility in a given area using Biogeography Based Optimization and Case based reasoning retrieval.

### 2.1.1 Modified approach to Biogeography Based Optimization

Biogeography is the study of geographic distribution of biological organisms and species. It is analogous to general problem solutions. BBO is an application of biogeography to optimization problems. It is modeled after the emigration and immigration of species between the islands. In the original BBO [Dan Simon (2008)], the author has described the

migration of the species from habitats with high HSI to the habitats with low HSI. In our proposed work, we have taken the CBR case base that is consisting of cases with their solution classes. These solution classes reveals the groundwater possibility i.e. High, moderate and low possibility. The cases are assumed to be the species that are residing in these habitats and their attributes are assumed to be the features of species. As the solutions are analogous to the habitats [Dan Simon (2008)], Case base is classified into three habitats: High habitat, moderate habitat and low habitat. High habitat consists of species having high groundwater possibility. Similarly Moderate habitat consists of species having moderate groundwater possibility. Low habitat consists of species having low ground water possibility. These habitats are forming the search space from where the similar cases to the input species are retrieved. Assume an input case (case without solution class) as an input species that is to be placed into an isolated habitat. High, Moderate and Low habitats tend to have high HSI. They are analogous to good solutions. Isolated habitat tends to have low HSI. Emigration takes places from high HSI habitats to low HSI habitat. We have defined the HSI as Similarity threshold i.e. maximum number of features matched between an input species and stored species in case base. Thus the less valuable species are filtered out. This reduces the search space.

We have considered the following BBO parameters:

HSI = Similarity threshold

E= Maximum Emigration Rate

I = Maximum Immigration Rate

$S_{max}$  = Maximum number of species

$S_{min}$  = Minimum number of species

For Low Habitat,

$E_l$  = Number of species in Low Habitat

$I_l = 0$

For Moderate Habitat,

$E_m$  = Number of species in Moderate Habitat

$I_m = 0$

For High Habitat,

$E_h$  = Number of species in High Habitat

$I_h = 0$

For Isolated Habitat,  
 $E_i = 0$   
 $I_i = E_l + E_m + E_h$

$S_{max}$  = size of dataset  
 $S_{min} = 1$

### 2.1.2 Case based reasoning Retrieval

Retrieval is a major research area in CBR. The most commonly investigated retrieval techniques, by far, are the k-nearest neighbors (k-NN), decision trees, and their derivatives. These techniques involve developing a similarity metric that allows closeness (i.e., similarity) among cases to be measured.

We have used the typical nearest neighbor technique for case base retrieval [3]:

$$Sim(p,q) = \sum_{j=1}^n f(p_i,q_i) \times w_i$$

where,  $p$  and  $q$  are two cases compared for similarity

$n$  is the number of attributes in each case

$i$  is an individual attribute from 1 to  $n$

$w_i$  is the feature weight of attribute  $i$

Generally, the similarity calculation continues until all cases in the case library have been compared, and ranked according to their similarity to a target problem case. Similarities are usually normalized to fall within the range 0 to 1, where 1 means a perfect match and 0 indicates a total mismatch.

The proposed algorithm can be described with the following steps:

Step 1: Classification of the case base into habitats.

No. of Habitats = No. of Problem solutions

Step 2: Defining Isolated habitat in which the input case is placed.

Step 3: Defining Habitat suitability Index.

Assuming  $\square$  as the similarity threshold ( $\square$  is the number of features matched between an input species and stored species)

For each habitat

For each species

If (HSI = max ( $\square$ ))

Then Emigrate species that are most similar to an input species to the Isolated habitat

End

End

Step 4: Less valuable species are filtered out. Now, applying Case retrieval method on isolated habitat with fewer species to retrieve the best match to an input species from the retrieved species.

Step 5: Solution class of a best retrieved species is

## 3. GROUNDWATER POSSIBILITY DETECTION

### 3.1 Dataset

The dataset used in our experiment consists of groundwater observations made by domain experts to predict the groundwater possibility

Table 1: Six attributes used in dataset

Attributes Values

<b>Geology</b>	Sedimentary, Younger alluvium, Older alluvium, Igneous, Metamorphic
<b>Landform</b>	Floodplain, Intermontane valley, Pediment, Alluvial fans, Bajada, Pediplain, Buried pediment, Alluvial Plain, Deltaic Plain, Wadi, River terraces, Old meander etc.
<b>Soil</b>	Sandy loam, Sandy gravel, Coarse sand, Clay loam, Alluvial sand, Gravel sand, Gravel sand pebbles, Sand, Rocky etc.
<b>Land use</b>	Agricultural land, Forest, Cultivated land, Fallowland, Waterbody, Wasteland, Swampy land, Buildup, Urban, Grass, Shrubs, mixed vegetation etc.
<b>Slope</b>	Gentle, Steep
<b>Lineament</b>	Absent, Present

First column of table 1 gives the categorical attributes that are used in the dataset and the second column indicates the attribute values.

### 3.2 User Interface

The development of Graphical User Interface for the Groundwater Possibility Retrieval System is accomplished using Matlab ver 7.0. User is required to enter all the parameters in order to predict the ground water possibility of a given area.

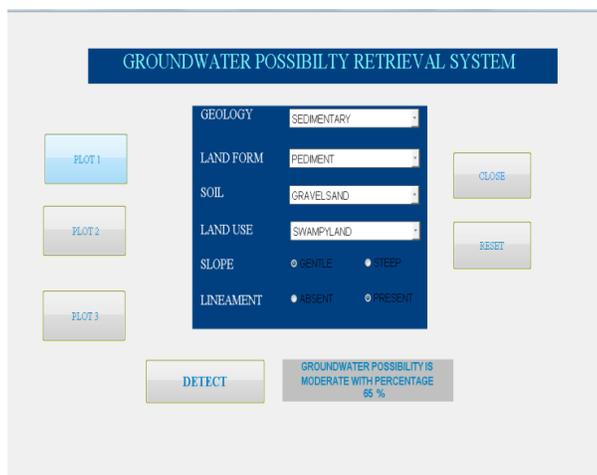


Fig1. Groundwater Possibility Retrieval system

### 3.1 Results

In GUI, the user has given the values to the parameters to detect groundwater possibility.

Example Query entered:

Table 2: Query Table

Geology	Younger Alluvium
Landform	Intermontane valley
Soil	Coarse Sand
Land Use	Cultivated land
Slope	Gentle
Lineament	Present
Groundwater possibility	?

Our aim is to find out the Groundwater possibility of the Query entered. Using Biogeography Based Optimization, Emigrate the species that are most similar to the query species from three habitats to the isolated habitat on the basis of Habitat Suitability Index.

In the graphs, we are showing the emigration of species. On X axis, Habitats are taken where 1 represents low habitat, 2 represents moderate habitat and 3 represents high habitat. The species which reside in low, moderate and high habitats respectively are represented by asterisk (\*). On Y axis, Number of species is taken. They are plotted on the graphs according to their index values.

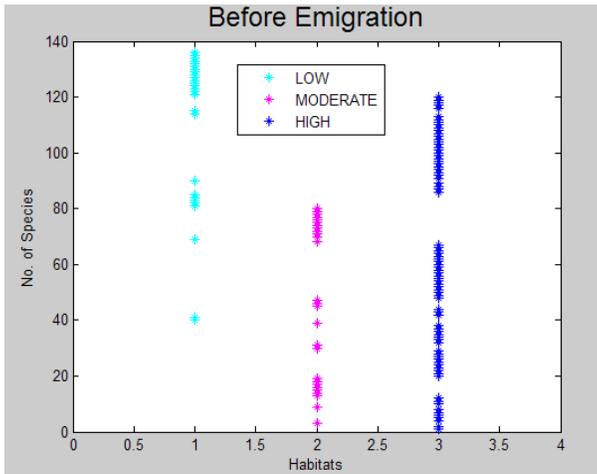


Fig2. Shows the number of species in low, moderate and high habitats before emigration.

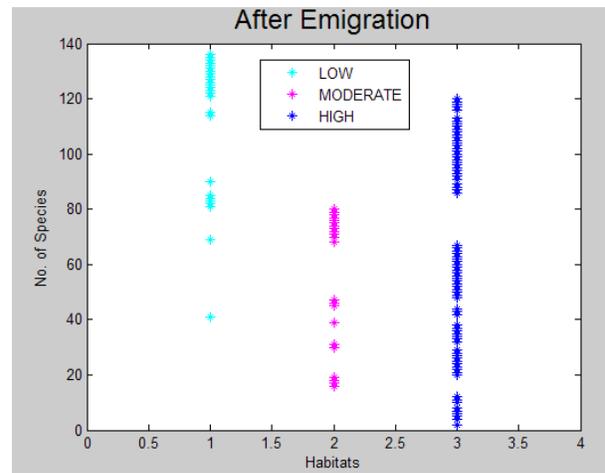


Fig4. Shows the remaining species in low, moderate and high after emigration

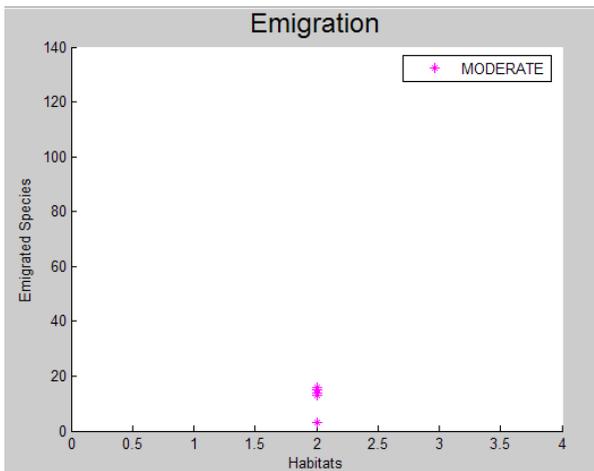


Fig3. Shows the emigrated species. Five species emigrated from moderate. No species from low and high habitats found similar to an input species.

These five species emigrate from moderate habitat to isolated habitat. Now by applying case retrieval method in isolated habitat, we retrieve a best matched species to the query species. Thus the solution of a retrieved species is transferred to the query species.

RESULT OBTAINED: GROUNDWATER POSSIBILITY IS MODERATE WITH PERCENTAGE 65%

#### 4. CONCLUSION AND FUTURE WORK

Based on our results presented in the previous chapter, We conclude that querying a large case base of groundwater possibilities which are similar to a present case i.e. query case, using Biogeography Based Optimization technique (application of biogeography to optimization problems) and case retrieval method (method to measure similarity among cases), can easily predict the groundwater possibility in a given area. Case retrieval method is an intelligent method which involves retrieving the most similar previous cases to provide a solution of a new decision problem. The Case base is designed with the help of a groundwater monitoring experts. Using the system, it is now possible to undertake a regional hydro geological survey to map the distribution of groundwater reserves both in plan and in depth with high efficiency, thus ensuring a reliable choice of the most productive areas for water supply bores. The system has been developed which is capable of determining groundwater content at various depths without drilling wells. The system developed is of great significance as the presence or absence of ground water in a given area has a direct effect on its real estate values. So the system can play a crucial role in economic sector of a nation. Moreover, it is very useful in military applications during the time of combats to find groundwater possibility in inaccessible areas like areas across the border of a nation. Movement of troops depends largely on such information as water is a basic need for survival.

However, one difficult part of this research is to find sufficient amount of cases. Most of the domain experts are not willing to provide their own know-how since this may cause losing their jobs. But, fewer cases will result in poor accuracy. If the mechanism of this study can be applied in practice to deal with more than thousand cases, the efficiency can be improved.

Future work can be done by experimenting with other categorical datasets to explore the fruitful results that can be applicable in other fields.

## REFERENCES

- [1] A. Wallace, “*The Geographical Distribution of Animals*”, Adament Media Corporation, 2006 (first published in 1876)
- [2] C. Darwin , “*The origin of species*. Gramercy”,1995(first published in 1859)
- [3] Dan Simon, “Biogeography-based optimization,” *IEEE Transactions on Evolutionary Computation*, vol. 12, pp. 702-713, December 2008.
- [4] T. Wesche, G. Goertler and W. Hubert, “Modified habitat suitability index model for brown trout in southeastern Wyoming”, *North Amer. J. Fisheries Manage.*, vol.7,pp. 232-237,1987
- [5] Sankar K pal and Simon C K Shiu, “*Foundations of Soft case based reasoning*”, Wiley Series on Intelligent systems.
- [6] Agnar Aamodt and Enric Plaza,”Foundational Issues, Methodological Variations, and System Approaches”, *Artificial Intelligence Communications*, IOS Press, Vol. 7: 1, pp. 39-59.