CHAPTER 2

LITERATURE REVIEW

Chapter 2

Literature review

2.1 Introduction

Understanding of how the HEA works are very crucial for this study. It is also very important to have the knowledge of all the water quality parameters considered in this study so that analysis and correct justification can be made wisely.

In this chapter, we will go back to basic where more detailed information regarding the subject matter will be discussed in addition to the explanation of relevant findings on few research papers that is ought to be similar.

2.2 Evolutionary Algorithm

Enrique Alba and Carlos Cotta define evolutionary algorithms (EA) as stochastic optimization techniques based on the principles of natural evolution. There is numerous of optimization techniques which mimics the functioning of biological processed.

Evolution is indeed a very complex process. Scientists such as Charles Darwin and Gregor Mendel have different theories when it comes to explaining evolution and its processes.

However, their theories can be generalized as follows:

1. Chromosomes are organic tools which encompasses the codes that store all the structure and features of living things. Evolutionary process operates on these codes rather than directly on organisms.
2. Evolutionary process takes place during reproduction where chromosomes were passed from one generation to one generation. Most important during reproduction is

mutation and recombination. Mutation introduces variability in the gene pool whereas recombination introduces the exchange of genetic information among individuals.

3. Natural selection is the mechanism that relates chromosomes with the adequacy of the entities they represent, favouring the proliferation of effective, environment-adapted organisms, and conversely causing the extinction of lesser effective, non-adapted organisms.

2.2.1 Process of Evolutionary Algorithms

Iteratively, EA generates tentative solutions for a certain problems. Collection of problems for a specified population comprises of individuals with one or more chromosomes were manipulated by EA. The chromosomes of each individual correspond to a conceivable solution for the respected problems.

Figure 2.1 shows the whole process of the algorithm. At the beginning of the process, population were generated randomly or by some heuristic procedures. Fitness values of each individual were measured to know how well the solution fits the problems in consideration. Consequently the fitness value will be used as search guidance within the algorithm.

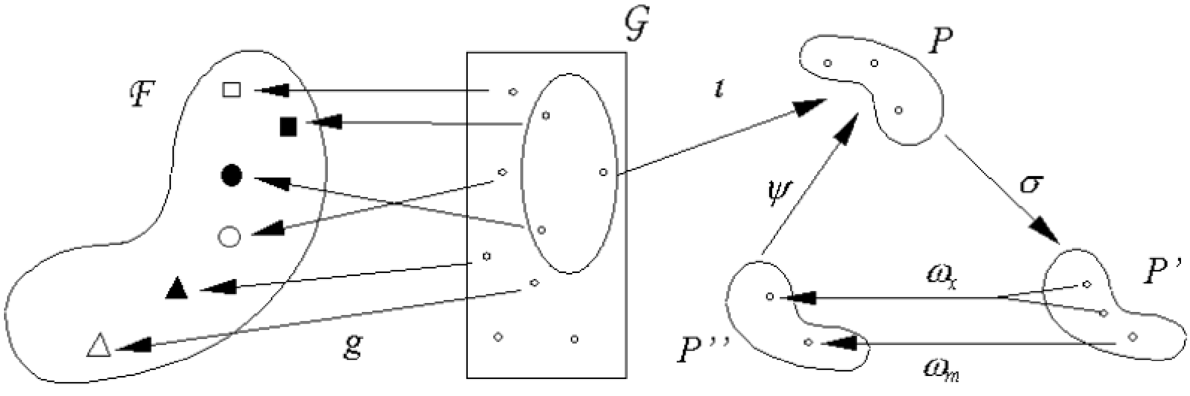

Figure 2.1: Evolutionary approach to optimization

2.2.2 Families of Evolutionary Algorithms

EA was known to have bloom in late 60's or early 70's. During this period, scientists from different part of the world have come up with different approaches to find the best and precise problem solving algorithms in the field. Existence of these different sources initiates the emergence of three different EA models. The primordial families are:

1. Evolutionary Programming (EP)

   Fogel *et. al.* introduces EP which focuses on the individual adaptation to the environment rather than manipulating their genetic information. Within this approach, evolutionary process was perceived on conceptual view where behaviour of individuals is directly modified, not their genes. Figure 2.2 shows the typical EP model using complex data structures such as finite automata or as graphs.

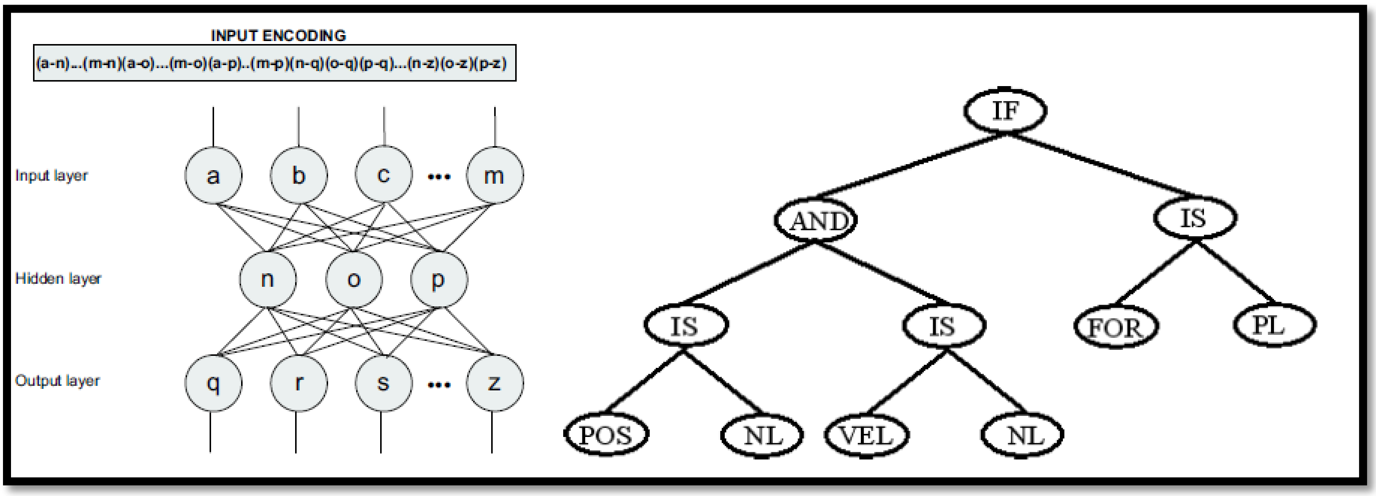

Figure 2.2: Examples of complex representations. (Left) Graph representing neural network. (Right) A tree representing a fuzzy rule.

2. Evolution Strategies (ES)

   This technique was co-founded by Rechenberg and Schwefel from Germany. It uses mutation as the unique reproductive operator as well as recombination. A special feature of ES is their capability of self-adapting towards the controlling of the application of mutation. Search progresses are optimized not only by evolving the solutions of problems under interest but also evolving the parameters involved for mutating these solutions.

3. Genetic Algorithms (GA)

   GA is the most common approach of EA introduced by Holland. Primary feature of GA is the employment of recombination or crossover operator as main search tools. It assumes that solutions can be discovered on individual basis but will create better solutions when they are combined. Mutation act as a secondary background operator whose purpose is merely to add additional information to the population.

These families were not totally isolated from each other. They are interconnected in some way that allows into the emergence of new variants in EA families as follows:

1. Evolution Programs

   This terms is introduced by Michalewicz which comprises techniques such as GAs and EP by means that it uses the principles of functioning of GAs but in the same time evolve complex data structures as in EP.

2. Genetic Programming (GP)

   GP can be understood as an EP by which the structures evolved are represented as computer programs. John Koza, scientist who has been introducing this approach, aims at programming automated solution for certain task, formulated as a collection of I/O examples. These programs are typically encoded by trees. (See Figure 2.1 right)

3. Memetic Algorithms (MA)

   Generally, memetic algorithms is problem-aware EA. EA were combined with local search algorithms such as heuristic algorithms, hill-climbing or branch-and-bound to perform the problem-awareness features.

2.3 Hybrid Evolutionary Algorithms (HEA)

Hybrid Evolutionary Algorithms (HEA) is a technique which integrates different learning and adaptation techniques, to overcome individual EA limitations and achieve optimized problem solving solution through the hybridization of these techniques.

HEA combines genetic programming (GP) and genetic algorithms (GA) to generate and optimize the structure of rule sets and to optimize the parameter of a rule set respectively. GP is an extension of GA whereby computer programs of varying sizes and shapes acts as the initial genetic population.

Computer programs are represented as parse tress in standard GP and the element of a function set are represented by a branch node. These elements can comprise of logic operators, arithmetic operators, elementary functions etc. Elements from a terminal set such as variables, constants and functions are represented by leaf node.

Evaluation of this program is done by considering the “fitness cases”. Next generation are created from recombination of fitter programs by crossover and/or mutation or also known as genetic operators.

Iteration will continue repeated for generations and will stop the run after satisfactory termination criterion is achieved. The random parameters in the rule set will be optimized by general GA. Figure 2.3 shows the general flowchart of HEA processes.

2.3.1 Structure Optimization by GP

Suppose that we have a rule set as below:

IF (exp(SS)<48.180) .....T_{IF1}

THEN DO=pH .....T_{THEN1}

ELSE .....T_{ELSE1}

DO=ln(|((-169.904)-(ColiF*(NO3-0.007)))|) .....T_{THEN2}

It can be written in the vector of binary trees representation: (T_{IF1}, T_{THEN1}, T_{ELSE1}, T_{THEN2})

3 functions sets were defined as such:

Logic function set: FL = {AND, OR}

Comparison function set: FC = {>, <, ≥, ≤}

Arithmetic function set: FA = {+, -, *, /, sin, cos, exp, ln}

The IF_TREE can contain the combination of these 3 function set, and the THEN/ELSE_TREE can only comprise of arithmetic function set. To control the complexity of the rule set we have a predefined maximum size of a rule set (MAXK) and the maximum tree depth (D_{IF} and $D_{THEN/ELSE}$ for the IF_TREE and the THEN/ELSE_TREE respectively).

2 level of crossover are available for binary vector tree:

1. Vector-level crossover

2 parents, A and B were considered for crossover. A random position was selected as a crossover point and then we swapped the subsequent IF-THEN-ELSE statements below the point. After this process we have 2 new offspring rule sets. (See Figure 2.4).

2. Tree-level crossover

Same process as the vector-level crossover, suppose we have 2 parents, A and B. IF_Tree from each parent were randomly chosen as well as the crossover point were picked. Subtrees rooted at the crossover points were swapped producing 2 new offspring trees. One of them was selected as the corresponding IF_TREE as long as its maximum depth does not exceed D_{IF}. (See Figure 2.5).

There is also tree-level mutation which basically act the same only that swapping of tree nodes occur between the parents and a new subtree randomly generated producing the new offspring tree.

After all these processes, all individuals in every generation will undergo a process called simplification of rule set. This process helps to reduce the total number of parameters to optimize without changing the fitness of the rule set. This process should be done prior parameter optimization by the genetic algorithm.

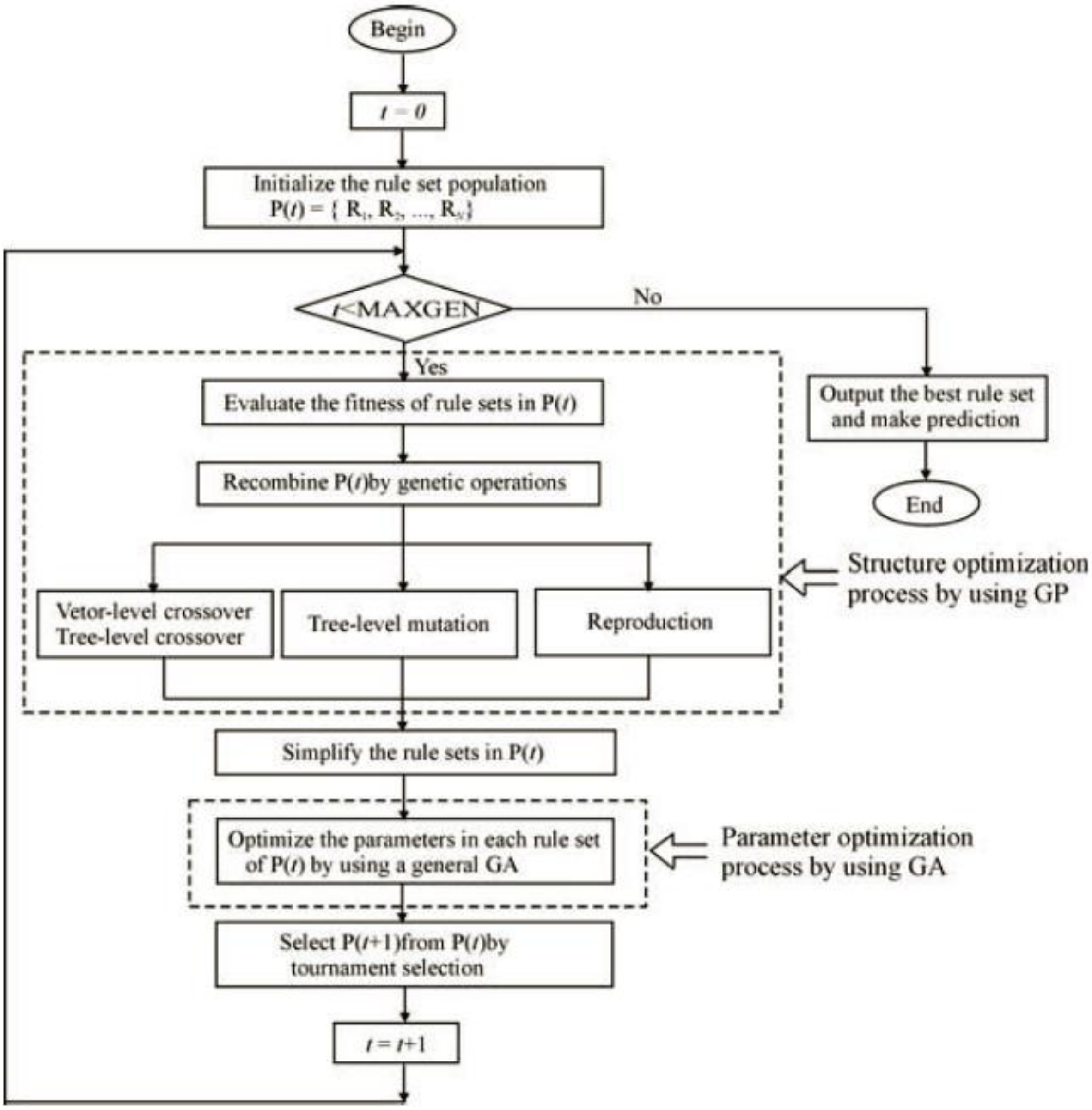

Figure 2.3: General flowchart of HEA.

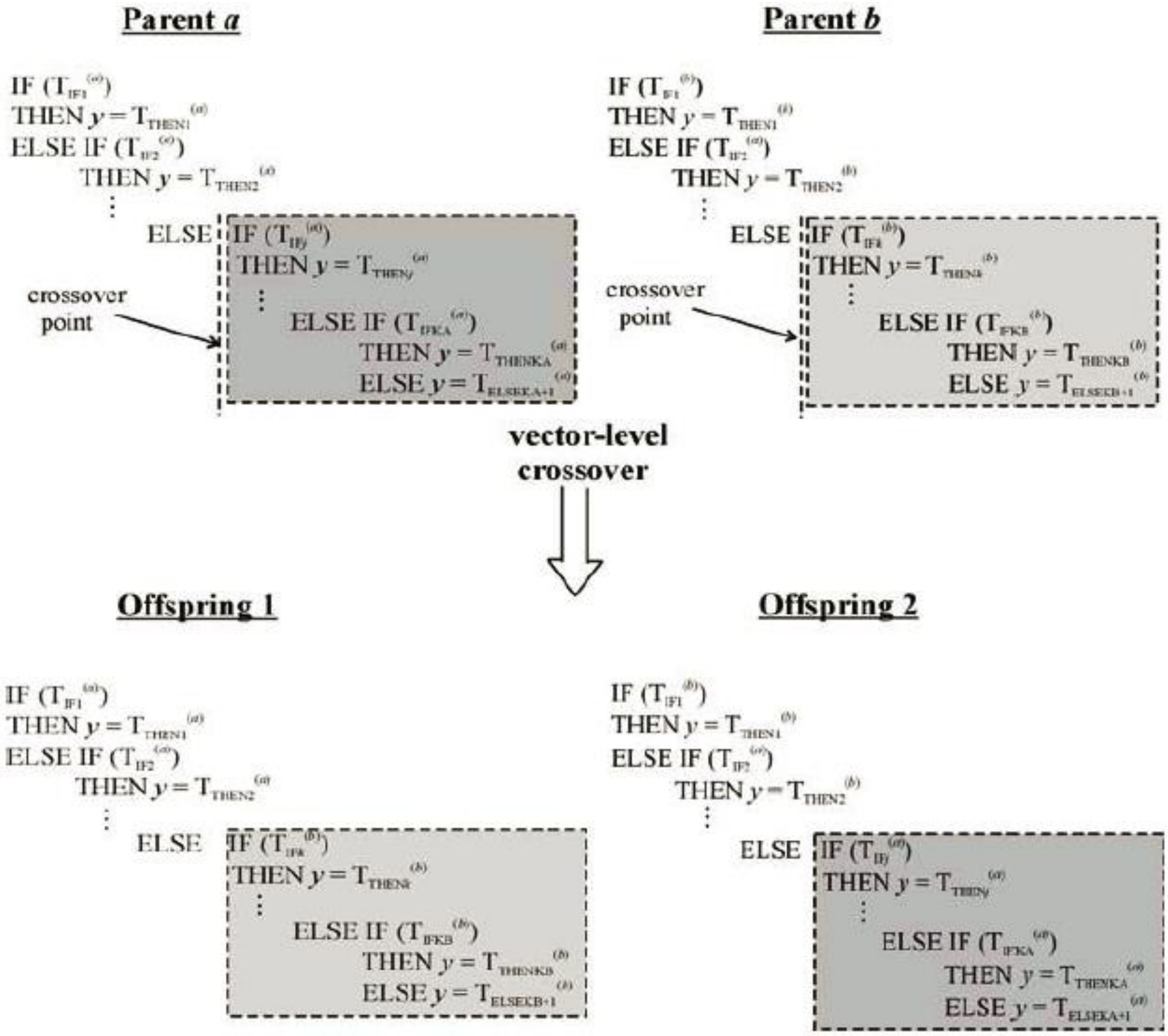

Figure 2.4: Example of vector-level crossover

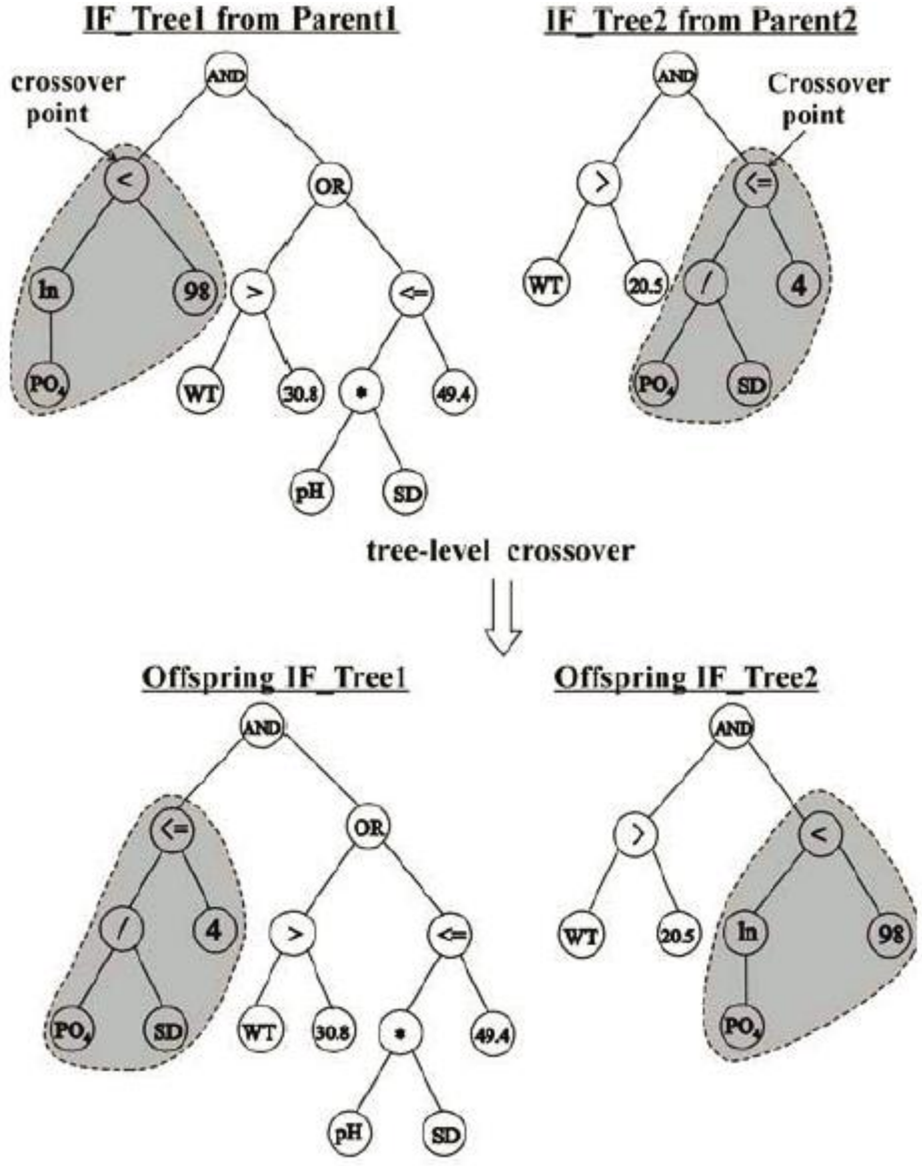

Figure 2.5: Example of tree-level crossover

2.3.2 Parameter Optimization by GA

Parameter of a rule set plays an important role to determine the accuracy rule set. Thus these parameter are ought to be optimized on each generation to generate a better problem solving solution for prediction of evolutionary data.

Parameter optimization by GA was done as follows:

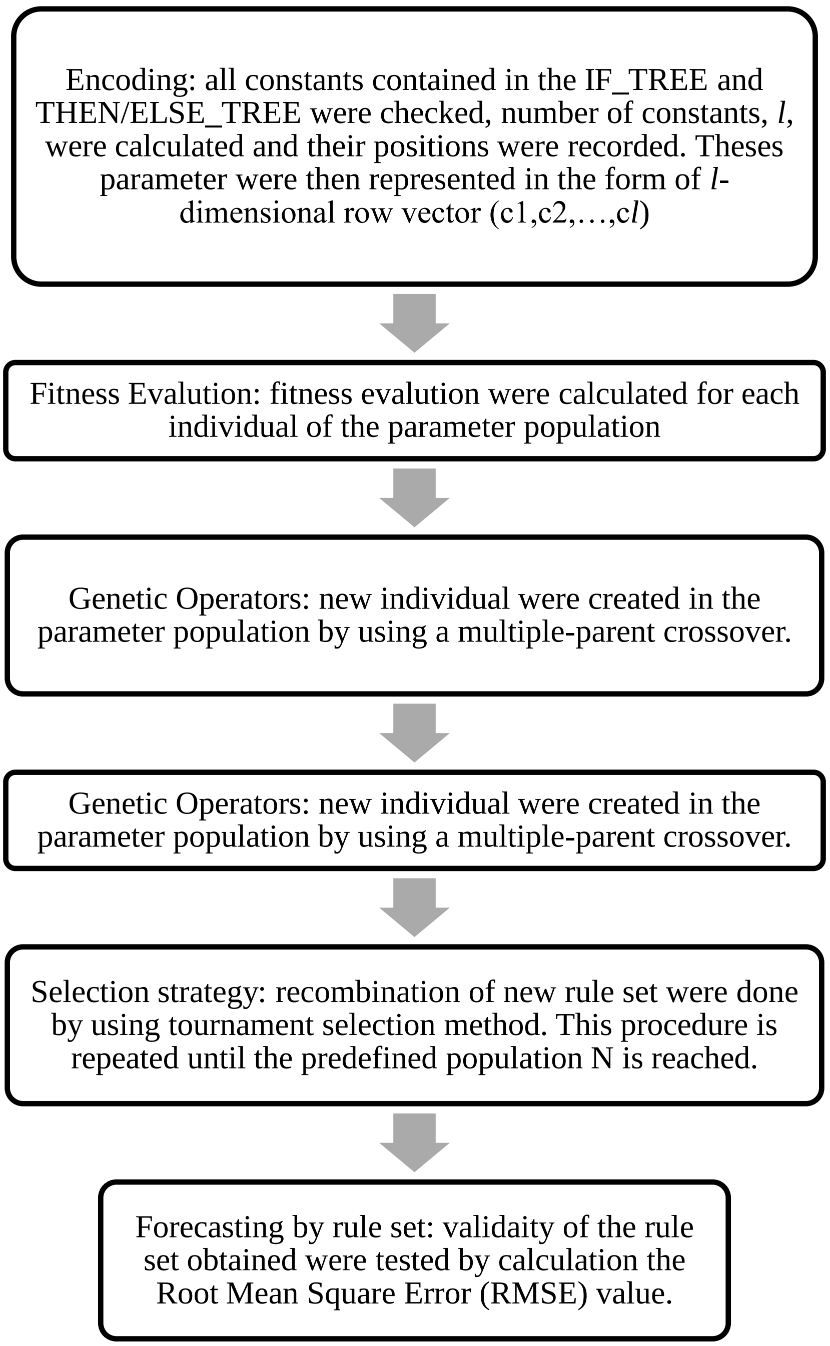

Figure 2.6: Process of Parameter Optimization by GA

2.4 Water Quality Parameters

2.4.1 Dissolved oxygen (DO)

Dissolved oxygen is a measure of the amount of oxygen that is dissolved or carried in a given medium. Dissolved oxygen probe are usually used to measure the amount of oxygen in a given medium, i.e. an oxygen sensor or an optode. Dissolved oxygen in water came from various sources or processes. Mainly it is by the diffusion from the surrounding air. Other than that, oxygen were produced from the water ventilation effect from water tumbling over falls and rapids. Photosynthesis also produces oxygen as a waste product.

Warm water cannot hold much oxygen within it. There will be more living things that may overpopulate such as the aquatic animal or bacteria within the area. At this point DO were used in great amount thus competency against it is increased causing the concentration of DO being reduced. Over-fertilization by run-off from farm fields which contains the phosphates and nitrates also contributes to the reduction of oxygen levels in water.

2.4.2 pH

The balance of positive hydrogen ions (H+) and negative hydroxide ions (OH-) in water determines how acidic or basic the water is. Concentration of H+ ions is balanced with OH- ions in pure water. Measures of pH at this point are 7 or neutral.

The age of the lake or pond will contribute to the pH value of the water bodies as well as the discharged chemicals from surrounding industries and communities. Usually lakes become acidic because of the organic materials build-up after certain of time since they're first formed.

2.4.3 Water temperature

The most common physical assessment of water quality is the measurement of temperature. Chemical and biological characteristics mostly dependant on the temperature of the surface water as temperature were able to change the composition of these characteristics. Dissolved oxygen is one characteristic that were directly affected by the change of temperature in the water bodies. Other than that, temperature also plays a role in determining the metabolic rates of aquatic organism, and how sensitive are these organisms to parasites, diseases and pollution, also the rate of photosynthesis of plants in the water.

Warm water is less capable of holding dissolved oxygen. For this reason, temperature should be measured at the same place within the stream at which dissolved oxygen is measured. This allows the correlation between the two parameters to be observed.

2.4.4 Salinity

Salinity is the measure of saltiness or dissolved salt content of a body of water. It is an ecological factor that can be considered important because it can influence the types of organisms that live in a particular water body. Also, salinity influences the kinds of plants that will grow either in a water body, or on land fed by water or by a groundwater. However, salt is difficult to remove from water, and salt content is an important factor in water use.

A plant adapted to saline conditions is called a halophyte. Organisms that can live in very salty conditions are classified as halophiles whereas an organism that can withstand a wide range of salinities is classified as euryhaline.

2.4.5 Turbidity

Turbidity is the optical property of a water sample that causes light to be scattered and absorbed rather than transmitted in straight lines through the sample. How much suspended material present in the water bodies will affect on light's ability to pass through the water. Plankton and soil erosion from human logging activities or mining are the most regular causes for turbidity.

The most accurate way to determine water's turbidity is with an electronic turbidimeter. The turbidimeter has a light source and a photoelectric cell that accurately measures the light scattered by suspended particles in a water sample. Turbidity can also be measured by filtering a water sample and comparing the filter's colour to a standard turbidity colour.

2.4.6 Ammonia (NH_3N)

Ammonia is a colourless gas with a strong pungent odour. It is easily liquefied and solidified and is very soluble in water. In freshwater, the principal sources of ammonia are; excretion by fish and other livestock as a normal part of their metabolism, the breakdown of protein in uneaten food or dead livestock which remains undetected and areas of biological filters deprived of oxygen may break nitrates down to ammonia.

As ammonia is released into the water by either of these processes it may take one of two forms:

1) **Free ammonia** (unionized ammonia, NH_3). This form of ammonia is highly toxic to fish.
2) **Ammonium** (ionized ammonia, NH_4+). This form of ammonia is virtually non-toxic to fish.

NH_3 is the principal form of toxic ammonia. It has been reported toxic to fresh water organisms at concentrations ranging from 0.53 to 22.8 mg/L. Toxic levels are depending on temperature and pH of the water. Toxicity negatively correlated with both pH and temperature as it gets decreases when pH and temperature are high. Plants are more tolerant of ammonia than animals, and invertebrates are more tolerant than fish.

2.4.7 Nitrate (NO_3^-)

Nitrate is the end product of the break-down of ammonia. Thus its presence indicates that the filter is functioning. In many cases the absence of nitrate is more worrying, as it may indicate that the filter is not functioning.

Nitrate is generally of low toxicity though some species, especially marines, are sensitive to its presence. Unlike temperature and dissolved oxygen, the presence of nitrates does not have a direct effect on aquatic insects or fish. However, excess levels of nitrates in water can create conditions that make it difficult for aquatic insects or fish to survive. When nitrate levels are high, as a result of biological filtration, other toxic chemicals produced in.

Algae and other plants use nitrates as a source of food. If algae have an unlimited source of nitrates, their growth is unchecked. Large amounts of algae can cause extreme fluctuations in dissolve oxygen.

2.4.8 Biological Oxygen Demand (BOD)

Biological Oxygen Demand (BOD) is a measure of the oxygen used by microorganisms to decompose dead organic matter. Large quantity of organic waste in water supply will cause the presence of bacteria, thus the demand for oxygen will increase causing the high level of BOD. As the waste is consumed or dispersed through the water, BOD levels will begin to decline. Nitrates and phosphates in a body of water can contribute to high BOD levels. At high BOD levels, organisms such as macroinvertebrates that are more tolerant of lower dissolved oxygen may appear and become numerous.

2.4.9 Chemical Oxygen Demand (COD)

Chemical Oxygen Demand (COD) is a measure of the capacity of water to consume oxygen during the decomposition of organic matter and the oxidation of inorganic chemicals such as ammonia and nitrite. It indicates the amount of oxygen consumed to completely chemically oxidize the organic water constituents to inorganic end products.

COD is related to BOD but it is important to understand that COD and BOD do not necessarily measure the same types of oxygen consumption. The BOD test measures the oxygen demand of biodegradable pollutants whereas the COD test measures the oxygen demand of biodegradable pollutants plus the oxygen demand of non-biodegradable oxidizable pollutants.

2.4.10 Chlorophyll-*a*

Green pigment that are usually in plants are known as Chlorophyll-*a*. Chlorophyll-*a* undergo photosynthesis; converts the sunlight into sugar in the process. Chlorophyll-*a* concentrations are usually used as indicator to measure the abundance of phytoplankton and biomass existed in the water bodies. Chlorophyll-*a* are also used to effectively measure the maximum photosynthetic rate activities occurring in the water bodies, assessing current tropic status and best applied for water quality measurement. High concentration of chlorophyll-*a* means a poor level water quality meanwhile low concentration of chlorophyll-*a* often suggest good water quality.

2.4.11 Conductivity

Conductivity is a measure on the ability of water to pass an electrical current. Inorganic dissolved solids existence in water such as chloride; sulphate, sodium and calcium affect the water conductivity by the chemical reaction of the ionized ions.

For stream and rivers, conductivity is affected by the geology of the area which water flows. The lower the conductivity means the better the water quality is as it indicates lesser dissolved solids presented in the water bodies. Higher readings means the water is highly polluted from nearby industrial activities or residential waste or other potential source. Long dry periods and little water flow can also contribute to higher conductance readings.