CHAPTER 2 : LITERATURE REVIEW

2.1 Water Studies and Statistics

Development of statistical analysis to water studies

Application of "statistics" is widely and extremely important to scientific experiment as a tool for experiment plan, data record, estimation and conclusion. Beveridge gave the importance of statistic to scientific researches that "Some knowledge of statistical methods is necessary for any form of experiment or observational research where numbers are involved especially for the complex experiments where there are more than one variable".

The role of statistics in water studies and water quality management has been expanding rapidly in recent years as its importance in providing information for a better understanding and prediction of the behavior of water quality variables in the environment and a water quality management decision-making. The use of statistical procedures with corresponding software are now as a part of the water quality monitoring system (Ward and Loftis, 1989).

The advanced statistical procedures have been widely used in various studies. Factor analysis, one of advanced statistical analysis was mentioned by Pisarskii and Khaustov (Pisarskii and Khaustov, 1982) in their researches as the most effective method to solve the problems when studying the conditions and processes of formation of groundwater runoff with a large amount of features (characteristics) which are not amenable to direct measurement. This method was used for constructing a factor model of formation of groundwater runoff. In water quality assessment of the Chao Phraya River summarized by Lohani (Lohani et. al., 1984), the statistical approaches to water quality evaluation were performed by using various methods including factor analysis. These mentioned that factor analysis provided a rapid water quality evaluation and being possible to reduce the number of variables measures and identify causes and sources of pollution.

Apart from factor analysis, regression especially multiple regression analysis is also widely used in water quality modeling for long term of the water quality management. Drew (Drew, 1983) applied the multiple regression analysis to establish whether there were any strong correlation between catchment factors of Lake Eppalock and specific quality characteristics. The multiple regression analysis approach was proving quite useful, although the statistical interpretations were very difficult to be managed. The results of the study summarized that catchment and major stream characteristics are such as to suggest that controlled development must be practiced to preserve the lake in its current situation. In the application of water quality modeling, Nagy and Butters (1988) used regression to provide an overall approach for quantitative describing and predicting water quality parameters in Lake Burley Griffin.

Based on the results of long term monitoring program, two water quality models were developed to predict the response of the lake to eutrophication control measures. The models indicated that the planned reductions in phosphorus at the major point sources, Queanbeyan Sewage Treatment Works (STW), would result in significantly lower levels of blue-green algae in the lake. The models also indicated that algae numbers were influenced by other factors apart from phosphorus. The importance of these findings to Australian conditions described, and the applicability of the overall approach to other water quality management problems was indicated (Nagy and Butters, 1988).

In addition to the application of multiple regression analysis, the study of regional assessment of NLEAP NO_3-N leaching indices was done by Wylie, Shaffer and Hall at South Platte River in Northeastern Colorado. In the study, the results of regression analysis indicated that although inorganic and organic fertilizer scenarios had similar r^2 values, the Feedlot Indicator variable was significant over and above the NO_3-N leached index. The analysis also showed that combination of either Movement Risk Index (MRI) or NO_3-N concentration of the leachate with the NO_3-N leached index leads to an improved regression, which provided insight into area-wide associations between agricultural activities and groundwater NO_3-N concentration (Wylie, Shaffer, and Hall, 1995).

For the study of water quality classification, "cluster analysis" has been applied in order to group the most similar entities together, for example, the study of characterization of the water quality in the Saint Lawrence River by Lachanee and Bobee (Lachanee and Bobee, 1979). A cluster analysis applied to the results of correspondence analysis leads to the definition of homogeneous zones which were characterized by values of the three parameters considered; nitrogen, total phosphorus, and turbidity. The results entered into the design processes of quality network for the Saint Lawrence River by allowing for the elimination of a certain number of station. The results showed the hierarchical tree obtained by discriminating a maximum of 13 groups of stations and 7 groups were obtained. Suwanrat and Watchawong (Suwanrat, 1996 and Watchawong, 1996) studied the classification of different running water systems in Chiang Mai based on their physico-chemical properties and benthos macroinvertebrate composition by applying "cluster analysis" and "matrices". Results from cluster analysis (dendograms) as well as the total matrix scores in both dry and wet seasons indicated the major differences between sewage canal sites and the other sites according to their high BOD_5, alkalinity, conductivity and low DO and total matrix scores.

The physico-chemical variables of water quality could be affected by biological organisms locally situated in aquatic ecosystems. Aquatic macroinvertebrate faunas (Benthos) are often used for biological monitoring due to their diversity, long-lived, often sedentary species that react strongly and predictably to human influences on aquatic ecosystems (Rosenberg and Resh, 1993).

Many statistical procedures have been included such as Detrended Correspondence Analysis (DECORANA), Two-Way Indicator Species Analysis (TWINSPAN), Multiple Discriminant Analysis (MDA), etc. A study of preliminary classification of running water sites in Great Britain based on macroinvertebrates and the prediction of community type using environmental data conducted by Wright (Wright et. al., 1984) presented that the sites were distinguished between type and length of a river. TWINSPAN and MDA have been found to be useful approaches to the classification of running water sites by their macroinvertebrates and the prediction of community (as indicated by the occurrence of species in the sites comprising the group) using 28 environmental variables (Wright et al., 1984).

In earlier years, Ormerod and Edward had studied the ordination and classification of macroinvertebrate assemblages in the catchment of the river Wye using the same statistical procedures in order to establish predictive relationship between benthos and physio-chemistry of running waters. They found out that the 45 sites were correlated with pH or total hardness and with slope or distance from source. The assemblages at soft-water sites were composed mostly of Plecoptera but at hard-water sites, even at high slope, the fauna was dominated numerically by Ephemeroptera. (Ormerod and Edwards, 1987). It is well known that base-poor acidic streams in temperate areas have impoverished macroinvertebrate fauna. Moreover, studies in several geographical areas have classified invertebrates using TWINSPAN and related site groups to stream acidity. Macroinverbetrates of 104 sites in upland Wales were ordinated by DECORANA, classified by TWINSPAN and related to stream chemistry by product-moment correlation. A similar pattern was apparent, in which pH and aluminum concentrations were the variables most closely related to the ordination, classification, and taxa richness of macroinvertebrates (Wade et al., 1989).

In 1990, the relationship between the physicochemistry and macroinvertebrate faunas of British upland streams and the development of modeling and indicator systems, for predicting fauna and detecting acidity were studied by Rutt et.al. (Rutt, Weatherley, and Ormerod, 1990).

The correlation of DECORANA axes as well as MDA indicated that sets of environmental variables, which principally included pH or aluminum and calcium concentrations, discriminated effectively between the stream groups defined by TWINSPAN

2.2 Water Quality Indices

History

The need for assessing water quality was first presented in Germany in 1850, as stated in Ott (1978). Studies were aimed at making the connection between the degree of purity of water and the presence of certain organisms.

Several groups of various countries gradually followed the trend, by using different classification systems to assess the water quality under their jurisdiction. There were two types of systems : those pertaining to the pollution level, and those dealing with micro-organisms and macro-organisms. These two classification systems were used to characterize aquatic environments at various pollution levels (Couillard and Lefebure, 1985). Opposite to this method, an indexing system for rating water quality in term of numeric scale was first developed in 1965 by Horton as a theoretical replacement to purely subjective methods of water quality classification. Since that time the ideas of Horton had been developed and applied quite extensively in the United States (Brown et al., 1970, 1972 ; Harkins, 1974) and in a much more limited way in Europe and the United Kingdom (Prati et al., 1971).

The Scottish Development Department (SDD) index had been most commonly adopted for the monitoring of surface water quality by the River Purification Boards (RPBs) in Scotland in 1976. The index was developed to improve existing methods of river quality classification. The efficiency of SDD index in evaluation surface water quality has been independently investigated by water authorities. Although in each instance the index scores produced agree favorable with subjectively ascribed river quality classification, a number of significant reservations were expressed (House and Ellis, 1987). The main criticism of the SDD (1976) index was that it has a tendency to underestimate at the lower end of the quality scale.

Water Quality Index Development

In the development of water quality index, most indices use parameters, weightings, rating curves and aggregation methods (Couillard and Lefebvre, 1985).

1. Selection of parameters (determinants)

The first stage in developing the index system was to assess which parameters provided fairly definite evidence of extraneous pollution. There are two methods prevalent used to select the determinants. Firstly the most relevant parameters can be selected from an agreement of experts (Ross, 1977) (House and Ellis, 1980).

Secondly, determinants derived from the use of statistical techniques such as regression (Harkins, 1974) or factor analysis (Joung et. al., 1979).

2. Weighting

The weighting aims to assign a relative importance that differs for every parameter. The relative importance is usually expressed through a coefficient, called weighting factor, and interrelates the importance of one given parameters with that of the various parameters used in the index. The sum of all weighting factors is generally 1.0. This way, the most important parameters are given the high relative weighting and the less important parameters are given lower weighting.

3. Rating Curve

The rating curves or subindex curves are used to transform parameter's concentration to the subindex scores. The x-axis represents the expected range of determinand values. The y-axis ranges from zero to some power of 10 (0-10, 0-100). This is the suitability-for-use axis (Smith, 1990). The rating curves of each parameters were constructed by refereed to a generally accepted standards and criteria or by mathematical expression (Prati et al., 1971).

4. Aggregation Process

The aggregation process is used to consolidate all quality scores of rating curves and, if necessary to weight those scores in terms of a given weight. It is after this step that the final result (water quality index) can be obtained. There are many aggregation functions developed.

Some aggregation methods are based on complex statistical considerations, while others use a combination of the additive form and simple parameters (Couillard and Lefebure, 1985) (Landwehr and Deininger, 1976).

In 1971, Prati, Pavanella and Pesarin had determined the mathematical expressions to transform determinant concentrations into levels of pollution expressed in new units which would be the units of measurement of pollutants and parameters. These had been drafted taking into account the rate of variation of the polluting effect with the increase in concentration of a given pollutant (Prati et al., 1971).

The index originally proposed by Brown et al., hereafter called the "arithmetic water quality index", which is of the form

$$WQIA = \sum_{i=1}^{9} w_i T_i (p_i) = \sum_{i=1}^{9} w_i q_i \quad \text{.....................(1)}$$

where p_i = measured value of the i^{th} parameter

T_i = quality rating transformation (curve) of the i^{th} parameter value , p_i, into a quality rating q_i, such that $T_i(p_i) = q_i$

w_i = relative weight of the ith parameter such that $\Sigma w_i = 1.0$

Nine parameters were used. It was mentioned that the index shown above had changed to a new index, a multiplicative index :

$$WQIM = \prod_{I=1}^{9} q_i^{wi} \qquad ....(2)$$

where qi and wi are as defined above (Landwehr, 1974 and Landwehr and Deininger, 1976).

The index construction suggests that each parameters is of a different weight or importance for a water quality station. The possibility that such weights were unnecessary in distinguishing between different quality situation was explored by the formulation of two additional indices.

These are the unweighted arithmetic index (WQIAU)

$$WQIAU = (1/9) \sum_{i=1}^{9} q_i \qquad ....(3)$$

and the unweighted multiplicative index :

$$WQIMU = \left(\prod_{i=1}^{9} q_i \right)^{1/9} \qquad ....(4)$$

Harkins argued that the objective of using water quality index developed by Brown et al. was not really objective because a panel of experts rated the water quality parameters to be used and many users felled that different panels gave different rating, thus destroying comparability and objectivity (Harkins, 1974). Therefore, he suggested the use of interesting application of Kendall's nonparametric classification procedure for the development of a water quality index. This particular application consists of computing the standardized distance each observation lies from a well-chosen control observation. The standardized distances or indexes thus computed optimally combine p-measurements in such a way that points most nearly alike have similar index value and these most different have dissimilar index values.

The index proposed by Harkins, which is based on Kendall's nonparametric multivariate ranking procedure, the standardized distance is :

$$S_n = \sum_{i=1}^{p} (R_i - R_c)^2 / \mathrm{Var}(R_i) \quad \text{.....................(5)}$$

and

$$\text{Variance } (R_i) = 1/12n \; [\,(n^3 - n) - \sum_{i=1}^{k} (t_k^3 - t_k)] \quad \text{.....................(6)}$$

where

$i = 1,2,\ldots,p$

p = the number of parameters being used

n = the number of observation plus the number of control points

k = the number of ties encountered, and

R_c = the rank of the control value

From the index mentioned above, the only hindrance in using this method is the fact that the numbers generated in one evaluation cannot be directly compared with those generated by a different run and different time.

Landwehr discussed that there was an error in the statement of Kendall's procedure. Harkins gave the rank variance for each parameter as described in equation (6). Ignoring the confusion caused by a doubly defined i subscript, the equation would be reduced the following form :

$$\text{Variance}(R_i) = 1/12n \; [(n^3-n) - k(t_k^3-t_k)] \quad \text{....................(7)}$$

Therefore, he assumed that Harkins intended to use the correct form as followed (Landwehr, 1974).

$$\text{Variance}(R_i) = 1/12n \; [(n^3-n) - \sum_{j=1}^{k} (t_j^3-t_j)] \quad \text{....................(8)}$$

According to the 1974 Annual Report of the Council on environmental quality, this index, termed an objective water quality index, is of current interest to the Environmental Protection Agency (EPA). Kendal's Index values (S_n) are calculated as follows.

$$S_n = \sum_{i=1}^{q} \frac{(R_{in} - R_{ic})^2}{Var_i} \qquad n = 1,2,\ldots,q \quad \text{....................(9)}$$

$$Var_i = 1/12q \; [(q^3-q) - \sum_{j=1}^{ki} (t_{ij}^3-t_{ij})] \quad \text{....................(10)}$$

where S_n = Harkins index value for the n^{th} water sample in the data set

q = total number of water samples in the particular data set under consideration

p = total number of parameters being used

R_{in} = rank of the n^{th} water sample, according to the value of the i^{th} parameter, compared to the values of that parameter among all of the q water samples.

R_{ic} = control value of the i^{th} parameter

Var_i = rank variance exhibited in the ordering of the i^{th} parameters values

t_{ij} = number of elements involved in the i^{th} ties encounter in ordering the values of the i^{th} parameters

k_i = total number of ties encounter in ordering the values of the i^{th} parameters

Thus many water quality indexes had been developed, the comparison between the developed indexes and existing classification in order to test for agreement is needed. A comparative study between the Scottish Development Department (SDD), water quality index and the National Water Council (NWC)/Thames Water Authority (TWA). The SDD developed their WQI as a part of an investigation into the improvement of existing classification. The index was based on the work of Brown et.al. in which ten parameters were selected by a panel of experts. Rating curves for each parameters were constructed with reference to general accepted standards and criteria and graphically expressed parameter concentrations on a scale of 0-100. In this comparative study, three mathematical formulae were chosen . The study concluded that the Geometric weighted showing high degree of agreement with the NWC and TWA (House and Ellis, 1980).

A part from many advantages of developing the water quality index to assess water quality, WQI aims to be maximum value to water quality managers. A classification system should not only category water quality but also provide information on possible economics and beneficial water use.

House and Ellis studied the details of the development of an alternative WQI structure which had been designed for the evaluation of the UK surface water quality, to be used in potable water supply where the intended use was for fisheries purposes, which utilizes EEC legislative criteria in its formulation. The study results could be regarded as justifying the structure and efficiency of the purpose WQI method. The high level of agreement between WQI and NWC classification would suggest that general water quality index is at least as good as existing methods of water quality assessment.

In addition, the index allowed the economic benefit of management strategies which are applied over fixed time scale to be evaluated in both quality and monetary terms. It was intended that the use of WQI might be further extended to form the basis of a cost-benefit analysis program for water quality management, thus further enhancing the application of best management practices in the water industry (House and Ellis, 1987) According to the development for operational management, four independent water quality indices were obtained. These indices consist of a general water quality index which related water quality to a range of possible water use, the potable water supply index (PWSI) which reflects the suitability of raw water and for use in potable water supply and two indices to toxicity, the Aquatic Toxicity (ATI) and Potable Sapidity Indices (PSI), which were based on toxic determinants (House, 1986).

However, the results from these investigations highlighted a number of potential advantages in using a water quality indexing system in the operational management of surface water quality. Tyson and House again studied the decision of applying these indices to the full range of water quality data available within the water data archieving system of a water authority, in this instance, North West Water authority (NWWA) one of the largest in England and Wales. Computerized data handling enables the index to produce valuable management information in a timely and efficient manner. As a consequence NWWA adopted the WQI for internal management and reporting purpose (Tyson and House, 1989).