FUZZY RULE BASED MODEL FOR ESTIMATING AGRICULTURAL DIFFUSE POLLUTION

Binoy Alias M[1]. and P.P.Mujumdar[2]

[1]*Research Scholar, (binoyma@civil.iisc.ernet.in) and* [2]*Associate Professor, Dept. Civil Engineering, Indian Institute of Science Bangalore 560 012, India. (pradeep@civil.iisc.ernet.in)*

ABSTRACT:
A method based on fuzzy concepts is employed in this paper to obtain reasonable estimates of diffuse pollution from agricultural runoff using limited available information. Annual average use of herbicide per unit area, extent of herbicide applied area and herbicide application season are considered as fuzzy input sets and observed herbicide concentration at the basin outlet as the fuzzy output set. Fuzzy rules are generated from the available data sets from which combined rule bases are formed. These rules are then used for mapping the input space to output space using a defuzzification procedure. This method learns from historical information. The method is demonstrated with two common herbicides, Atrazine and Metalochlore for the White River basin (US). The model is capable of generating monthly average values for the herbicide concentration with a reasonable accuracy.

KEY WORDS: Diffuse pollution; fuzzy method; limited information; transport mechanism; uncertainty; vagueness

INTRODUCTION

Agricultural systems are highly dependent on agrochemical inputs such as fertilizers, pesticides and herbicides. Diffuse pollution in the form of runoff generated from the agricultural lands is one of the major sources of quality degradation of surface water bodies. Quantum of the agrochemical transport from the source depends primarily on their loading per unit area, total area applied and application period. Other factors such as runoff, characteristics of basin, properties of pesticides, application practices etc. also influence the pesticide/herbicide transport. Pollutant transport mechanism from the source to the outlet of a watershed consists of several interlinked processes such as deposition, re-suspension, adsorption, chemical reaction, biological uptake, decay and percolation. The uncertainty at the source input and the known and unknown processes of the pollutant transport makes its estimation complex.

Pareira (1990), Crawford (2001) and Capel and Larson (2001) in their studies on pesticides/herbicides, identified the major factors that control the pollutant transport. Capel et al., (2001) estimated the annual pollutant transport as percent of use (load as percent of use - LAPU). Larson and Gilliom (2001) developed a regression model for pollutant estimation. The objective of this work is to address the uncertainty in the chemical input and transport processes through fuzzy concepts. A fuzzy relation between the influencing factors and the stream concentration at the basin outlet is identified from the observations for White River basin (US). Using the fuzzy relation, a fuzzy rule based model is developed to estimate the average monthly pollutant concentration at the outlet of the watershed (Hazeltone). The procedure is data driven and no physical and chemical processes are considered in the fuzzy rule based model.

METHOD

Fuzzy Inference System (FIS) is an effective tool to address uncertainty due to imprecision and vagueness, especially when the data availability is limited. The vagueness, subjectivity or imprecision in averaging the pesticide application rate, approximation of the applied agricultural area and the variation in pesticide application time, can be addressed effectively by the FIS method. The method also has the advantage of being extremely simple computationally.

The method developed by Wang and Mendel (1992) is used to construct a FIS for estimating the diffuse pollution due to agricultural runoff at the outlet of a watershed. For a known set of input-output data pairs: $(x_1^{(1)}, x_2^{(1)}; y_1^{(1)})$, $(x_1^{(2)}, x_2^{(2)}; y_1^{(2)})$,... where x_1, and x_2 are inputs and y is the output, the task is to generate the set of fuzzy rules and use that fuzzy rules to determine a mapping $f : (x_1, x_2) \quad y$. The procedure consists of following steps:

Step 1. Dividing the input and output spaces into fuzzy regions

To generate the fuzzy rules from the available information, the first step is to define the fuzzy regions in the input and output space, respectively. The possible domain interval of both the inputs and the output are divided into a number of regions in such a way that they overlap each other. The length of region may differ for each variable. One membership function is assigned to each region. The shape of the membership function can also vary.

Step 2. Generation of fuzzy rules from available data pairs

The data pairs $x_1^{(i)}, x_2^{(i)}; y_1^{(i)}$ are then to be assigned to the region where their membership value $(m(x_1), m(x_2); m(y))$ is the highest. The corresponding input and output regions can be denoted as I_i^{x1}, I_j^{x2} and O_y^k. These input-output region pairs constitute a rule. Then the fuzzy rule set is "IF x_1 is I_i^{x1} AND x_2 is I_j^{x2}, THEN y is O_y^k". The rules generated in this way are

'AND' rules. i.e. rules in which the conditions of the IF part must be met simultaneously in order for the result of 'THEN' part to occur. 'AND' rules are used here since antecedents are different components of a single input vector.

Step 3. Assigning a degree to each rule

When the rules are generated using all the available data pairs, it is highly probable that there will be conflicting rules. i.e. same premises (input regions) result in different consequences (output region). This conflicting situation is resolved by assigning a degree to each rule generated and accepting the rule which has the maximum degree. The degree of a rule is taken as the product of the membership values of input and output. For a rule, for example,
'IF x_1 is A AND x_2 is B, THEN y is C', the degree of rule = $m_A(x_1)m_B(x_2)_Cm(y)$.
This is important in practical application because real data have different reliability. For good data, higher degree (close to one) can be assigned and for 'wild data', lower degree (close to zero) can be assigned. Hence a human experience or expert knowledge may also be integrated into rule formation at this point.

Step 4 Construction of a combined fuzzy rule base

The form of a fuzzy rule base is illustrated through a matrix. The number of fuzzy regions in the input defines the size of matrix of combined fuzzy rule base. For a two input – one output case, a combined rule base can be formed with a two dimensional matrix as shown in Figure 1. For example, for a rule 'if x_1 is S_2 and x_2 is P_3 then y is O_8', the cell corresponding to the input region of x_1 and x_2 (here S_2 & P_3) can then be filled with output region from the observed data (O_8). Thus each cell in the matrix is assigned an output region included in the rule with the highest degree. When the data is limited, there can be some empty cells, for which, the information may not be available. In such cases intermediate empty cells are filled with output regions in an ascending or descending transition from the nearest rule fixed cell. Expert knowledge or belief is to be used for the purpose.

Figure 1 Fuzzy Rule Base

Step 5. Mapping from the input space to the output space using the rule base and a defuzzification

The rule base constructed as explained is considered as the fuzzy mapping of information to be obtained from the input data available. Output related to any input could be obtained using the rule base by a defuzzification procedure. Through the processes such as implication, aggregation and defuzzification using centroid method, a crisp value for concentration is obtained for each set of input.

The fuzzification – defuzzification procedure narrated here is data driven and is capable of learning from every new set of data. This makes the mapping able to update with every change in the course of time. Adjustment to length of region, alternate membership functions for input/output etc. are made to fine tune the output of the model. Different combinations of triangular, trapezoidal and Gaussian membership functions for input and output are tried for achieving better results. The best model is selected based on the least root mean square error (RMSE) between the model output and observed values.

APPLICATION

The procedure is applied to construct models for two most common herbicides, Atrazine and Metalochlore, in White River basin in Indiana (USA). Based on the studies by Pareira (1990), Crawfard (2001) and Capel and Larson(2001), the major influencing factors which contribute to the herbicide concentration at the watershed outlet are identified as its application rate (lb/Acre) and extent of area applied (percentage of cultivated area). A plot of mean concentration at basin outlet over a period of seven years (by the authors) indicates that pollutant concentration or its transport is high during the period just after the application and is not seriously dependent on the basin runoff. Crawfard also inferred this fact in his studies. Hence the herbicide application time is considered as the third influencing factor. Data on herbicide application statistics for the years 1994 – 2000 for the Indiana State is available from USDA. (http://www.nass.usda.gov/in/ind_agstat.html). Annual average pesticide application rate for Indiana State is used as first input for the model. Herbicides are generally applied to the corn and soybean during April-June in Indiana. This herbicide application season is used as second input. From the USDA data, the weighted average of herbicide application rates and percentage of area applied of the corn and soybean cropped area are worked out from USDA data and is given in Table 1. The percentage area applied is used as third input.
Water quality data have been collected by USGS over time in the study area. One to four observations were taken for various water quality parameters depending on season. Three to four observations during the pesticide/herbicide

application season and one to two observations during the other months are available. These data for the same period (1994 –2000) are collected (http://waterdata.usgs.gov/in/nwis/qwdata). Monthly mean values of the herbicide concentrations at the basin outlet are worked out from the available data and used as output. Initially, triangular membership functions with three classes are used for all variables. Then the number of classes is increased and lengths of the classes are altered till an acceptable accuracy is achieved.

**Table 1 Chemical Application Rate and Percentage Area Applied in Indiana State**

Year	Atrazine		Metalochlor	
	Application Rate (lb/Acre)	Weighted Percentage Area Applied	Application Rate (lb/Acre)	Weighted Percentage Area Applied
1994	1.35	87	1.89	31
1995	1.31	87	1.91	21
1996	1.31	91	1.84	20
1997	1.33	84	1.85	30
1998	1.36	89	2.04	25
1999	1.26	91	1.43	27
2000	1.41	80	1.49	23

For higher accuracy, the membership functions of the input and output variables are altered with triangular, trapezoidal, Gaussian and Gbell membership combinations. The root mean square error (RMS) between the observed and estimated value in each combination is worked out.

RESULTS & DISCUSSION

The model is trained with 6 years data (1994-1999) and tested for one year data (2000). Out of different membership functions combinations tried for inputs and output, Gbell membership gave the least error of 1.48ìgm/l for Atrazine and Triangular membership resulted in least error of 0.37ìgm/l for Metalochlor. Membership functions for the final selected model for both herbicides are shown in Figure 2(a) and 2(b).

A plot of the obseved data and estimated values obtained from the selected models are given in Figure 3. The predictive accuracy of the fuzzy model is very reasonable. Data scarcity for river basin influenced the estimation. From the very approximate data, the model is capable of generating reasonably accurate monthly average concentrations for both Atrazine and Metalochlore.
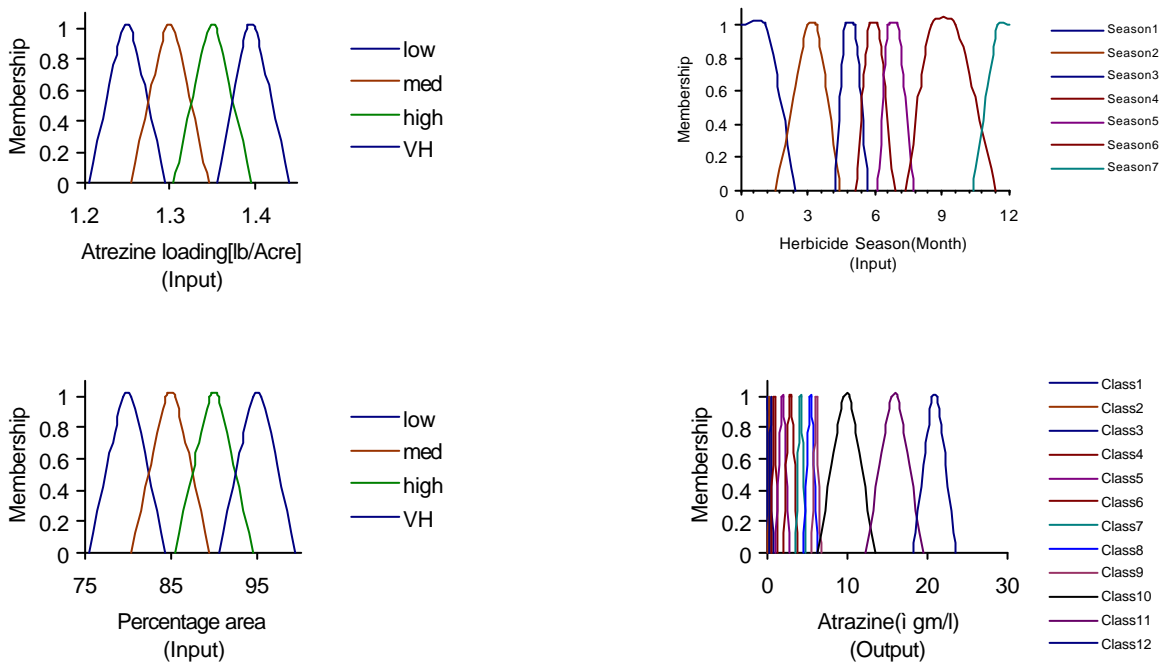
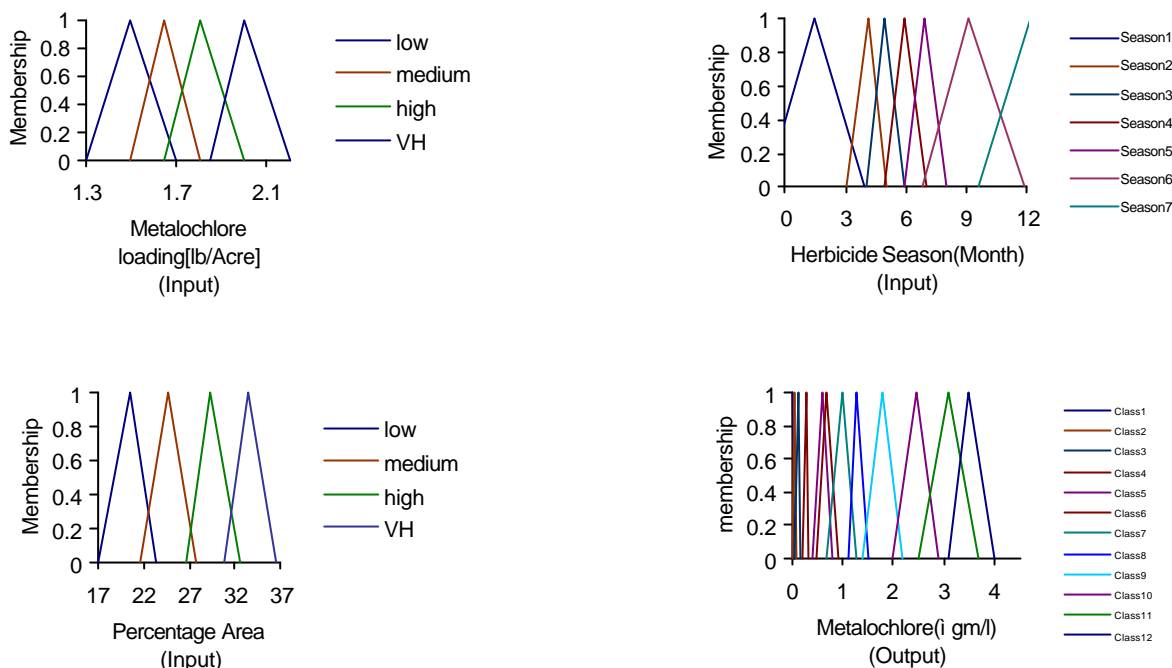**Figure 2(a) Membership Functions for input and output (Atrazine)**

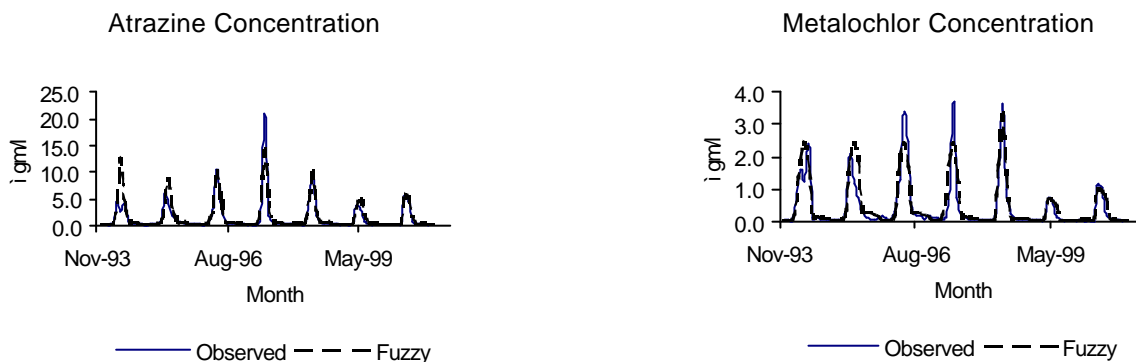*Figure 2(b) Membership Functions for input and output (Metalochlore)*

Figure 3 Plot of Observed and estimated values of Atrazine and Metalochlore

This method is applied for annual herbicide and nutrient transport for the same basin. The input used are average application rate, percentage applied area and annual runoff. The results are found to be very close with actual values when compared with regression models.

CONCLUSIONS

A fuzzy rule based model is developed for estimating monthly mean agriculture diffuse pollution concentration at the outlet of a river basin. Its capability is illustrated with two common herbicides, Atrazine and Metalochlore, for the White river basin in Indiana (US). The model works on general information like annual average herbicide use, percentage of cropped area in the basin to which the herbicide is applied and the time of application. Construction of the model is comparatively easy. The model is data driven and can be trained further with every new set of data. The linguistic structure of the model makes it transparent and easy to understand. The estimated values from the model are found to be reasonable and comparable with observed values.

Major limitation of the model is that it is site specific. Generalization can be made with wide range of data from different basins. Separate models are to be made for each pollutant. Accuracy of output depends on the reliability of the training data.

REFERENCES

Capel Paul D. and Larson Steven J. (2001), *'Effect of Scale on the Behavior of Atrazine in Surface Waters'*, Envron. Sci. Technol. Vol.35, No.4.pp 648-657

Capel Paul D. and Larson Steven J. and Winterstein Thomas A. (2001), *'The behavior of 39 pesticides in surface waters as a function of scale'*, Hydrological processes 15, pp1251-1269

Crawfard Charles G. (2001), *'Factors Affecting Pesticide Occurrence And Transport In A Large Mid-Western River Basin'*, Journal of American Water Res. Asstn.Vol. 37 No.1 pp1-15

Larson Steven J. and Gilliom Robert J. (2001), *'Regression Models For Estimating Herbicides Concentrations In U.S. Streams From Watershed Characteristics'*, Journal of American Water Res. Asstn.Vol. 37 No.5 pp1349-1367

Pareira E. Wilfred and Rostad Colleen E. (1990), *'Occurrence, Distribution and Transport of herbicides and their Degradation Products in the Lower Mississippi River and Its Tributaries'*, Envron. Sci. Technol. Vol.24, No.9.pp1400-1406

Wang Li-Xin and Mendel Jerry M. (1992), *'Generating Fuzzy rules by Learning from Examples'*, IEEE transactions on Systems, Man and Cybernetics, Vol. 22 No.6, pp1414–1427