13

Prediction of Herbicides Concentration in Streams

Raj Mohan Singh
Department of Civil Engineering, MNNIT Allahabad, India

1. Introduction

Natural and anthropogenic variables of stream drainage basins such as hydrogeologic parameters (permeability, porosity etc.), amount of agricultural chemicals applied, or percentage of land planted affect agricultural chemical concentration and mass transport in streams. The use of herbicides, pesticides, and other chemicals in agricultural fields increase the concentration of chemicals in streams which severely affects the health of human and environment. The transport of chemical pollutants into river or streams is not straight forward but complex function of applied chemicals and land use patterns in a given river or stream basin. The factors responsible for transport of chemicals may be considered as inputs and chemical concentration measurements in streams as outputs. Each of these inputs and outputs may contain measurement errors. Present work exploited characteristics of fuzzy sets to address uncertainties in inputs by incorporating overlapping membership functions for each of inputs even for limited data availability situations. Soft computing methods such as the fuzzy rule based and ANN (Artificial Neural Networks) is used for characterization of herbicides concentration in streams. The fuzzy c-means (FCM) algorithm is used for the optimization of membership functions of fuzzy rule based models for the estimation of diffuse pollution concentration in streams. The general methodology based on fuzzy, ANN and FCM for estimation of diffuse pollution in streams is presented. The application of the proposed methodology is illustrated with real data to estimate the diffuse pollution concentration in a stream system due to application of a typical herbicide, atrazine, in corn fields with limited data availability. Solution results establish that developed fuzzy rule base model with FCM outperform fuzzy or ANN and capable for the estimation of diffuse pollution concentration values in water matrices with sparse data situations.

Application of pesticides, insecticides and herbicides, cause diffuse pollution, commonly referred to as non-point source pollution in river or streams. Diffuse pollution from agricultural activities is a major cause of concern for the health of human and environment. Diffuse (non-dot, dispersed) pollution generally arises from land-use activities (urban and rural) that are dispersed across a catchment or subcatchment, where as point sources of pollution arise as a process industrial effluent, municipal sewage effluent, deep mine or farm effluent discharge (Novotny 2003, based on CIWEM (D'Arcy et al., 2000)). Potential point sources of pollution is characterised by its location, magnitude and duration of activity; and the sources of pollution is characterized when these parameters are identified

(Mahar and Dattta 2000; Singh and Datta, 2004, and Singh and Datta, 2006a and 2006b). In diffuse sources of pollution or non-point sources of pollution, sources of pollution is moving with polluting media thus making it more difficult and complex problem to solve.

Often diffuse pollution is individually minor but collectively constitutes significant sources at basin scale. Although nonpoint or diffuse sources may contribute many of the same kinds of pollutants, these pollutants are generated in different volumes, combinations, and concentrations (Jha et al., 2005). Thus, diffuse pollution comprises true non-point source pollution together with inputs from a multiplicity of minor point sources. The important characteristics of diffuse pollution are, therefore, not whether anyone can identify the source or sources, but the collective impact of diffuse pollutants and the mechanisms through which they move through the environment. The concept of diffuse pollution is useful because it explains features of pollution in receiving water bodies that differ from the point sources of pollution that are typically well characterized, monitored, and quantified. Some of the characteristics of diffuse pollutants are that the concentrations of some pollutants actually may increase with flow rather than it has diluted, pollution peaks are variable and difficult to predict, and impacts are often slow to develop and become evident years later (e.g. contamination of groundwater). For diffuse pollution, it is the proportion of the land use from which the pollution is derived, is more important.

Agricultural activities such as application of herbicides result in the contamination of surface water with agricultural chemicals. Numerous recent investigations (Goolsby and Battaglin, 1993 and 1995; Schottler et al., 1994; Baker and Richards, 1990) indicate that significant quantities of some herbicides are flushed from cropland to streams each spring and summer during rainfall events following the applications. Peak concentration of several herbicides can exceed 10 μg/l during these events (Coupe et al., 1995; Scribner et al., 1994). Pareira (1990), Crawfard (1995, 2001), Capel and Larson (2001), and Smith and Wheater (2004) in their studies on pesticides/herbicides, identified the major factors that control the pollutant transport. Herbicides and pesticides concentrations in surface waters are affected by natural and human factors. For example, concentrations of atrazine, a herbicide widely used on corn fields, tended to be higher in an agricultural basin with permeable, well drained soils, than in an agricultural basin with less permeable, more poorly drained soils (Crawfard, 1995). 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 the estimation of pollutants.

Water resources professionals, managers and government authorities involved in surface water management are increasingly pressed to make appropriate decisions on land use and development policies such that these decisions will not adversely affect the health and environment. At the same time, they are constrained by inadequate budgets, limited resources, and incomplete information, which compel them to rely on models to evaluate or to estimate the pollution characteristics in the water bodies, and the implications of their decisions based on those evaluations. In this regard, the role of complex stream quality simulation models e.g. SWAT (Arnold et al. 1983), AGNPS (Young et al., 1989) etc. in evaluating runoff pollution conditions under various agricultural chemicals and land use patterns is also limited. These models incorporate rainfall, catchments, and pollutant characteristics, requiring extensive calibration and verification. However, their results are not without large uncertainties. These uncertainties arise both in the representation of the physical, chemical, and biological processes as well as in the data acquisition and parameters for model algorithms. Consequently, the complexities of these models and their resource-intensive nature are significant obstacles to their application (Charbeneau and Barrett 1998).

There is a need for the development of simpler methods of agricultural stream quality predictions that provide the required information to the analyst and water managers with minimal effort and limited data requirements as compared to complex process models. As an alternative or supplement to complex runoff quality simulation models, fuzzy rule based model with FCM is proposed to estimate pollutant concentration due to applications of agricultural chemical, herbicide, atrazine, in the streams.

The herbicide atrazine (2-chloro-4-[ethylamino]-6-[isopropylamino]-1,3,5-triazine), a chlorinated herbicide, has been one of the most heavily used herbicides in the world. Atrazine is toxic to many living organism. The maximum contaminant level (MCL) of atrazine is restricted to 3 μg/l for drinking water (USEPA, 2001). Because atrazine is water soluble, it has the potential to leach into ground water and run off to surface water. Atrazine is associated with developmental effects (USEPA, 2002), such as birth defects, structural anomalies, and adverse hormone changes. Thus, its accurate estimation in water matrices is imperative.

In this study, a fuzzy rule based model optimized by fuzzy c-Means, is developed to obtain the estimate of atrazine concentrations from agricultural run-off using limited available information. The work discusses the methodology to develop the fuzzy rule base model using annual average use of herbicide atrazine per unit area, extent of herbicide atrazine applied area and herbicide atrazine application season as inputs to fuzzy rule based model and observed herbicide concentration at the basin outlet as the output for the fuzzy model. The data of White River Basin, a part of the Mississippi River system, USA, is used for developing the fuzzy rule base model.

2. Agricultural diffuse pollution concentration simulation in streams

Natural and anthropogenic variables of stream drainage basins such as hydrogeologic parameters (permeability, porosity etc.), amount of agricultural chemicals applied, or percentage of land planted affect agricultural chemical concentration and mass transport in streams. The general form of model that simulates the concentration measurement in a watershed can be represented by (Tesfamichael et al., 2005)

$$C = f\,(\mathbf{W}, \mathbf{H}, \mathbf{A}) \tag{1}$$

where C is the stream agricultural diffuse pollution observed concentration measurement values; **W** is a vector of watershed characteristics; and **H** is a vector of hydrological variables such as precipitation, runoff, etc., and **A** is a vector of relevant agricultural practices including actual chemical application rate in the field in lb/acre.

For a particular watershed, watershed characteristic, **W**, may be assumed to be constant. Also, for a particular hydrological unit, **H** may be assumed to be of similar characteristics. Then, Equation (1), though simplified, may be represented by

$$C = f\,(\mathbf{A}) \tag{2}$$

The **A** may be further represented by

$$\mathbf{A} = f\,(\mathbf{A}_C, \mathbf{A}_L) \tag{3}$$

where $\mathbf{A}_C$ represent the vector of applied agricultural chemical characteristics such as type of agricultural chemical (insecticide, herbicides etc.), application rate, application season etc., and $\mathbf{A}_L$ is the land use patterns such as type of crop grown, percentage of cropped area, etc.

Here, agricultural chemical considered is herbicide, atrazine, and crop considered is corn. In this study fuzzy rule based model with FCM simulates the stream system behavior from inputs of agricultural practices and corresponding observed concentration measurement values. In fact the model tries to emulate the mechanism that produced the data set. In this way, the mathematical description of the physical system is learned by the model, and therefore utilized as a tool for stream system simulation. The cluster centers of inputs and outputs obtained using FCM model, in essence, represents a typical characteristics of the system behaviour, and hence utilized in the formation of rule base of the fuzzy model.

3. Methodology

Statistical methodologies have been traditional being utilized for diffuse pollutants predictions in streams. However, transport of herbicides is complex and uncertain phenomena and traditional methods like regression are not able to incorporate uncertainty in model predictions. Present work will discuss methodologies based on recent soft computing techniques like fuzzy, artificial neural network (ANN) and their hybrids. The application of the proposed methodology is illustrated with real data to estimate the diffuse pollution concentration in a stream system due to application of a typical herbicide, atrazine, in corn fields with limited data availability.

3.1 Modeling approach

The models based on fuzzy logic and ANN, also known as intelligent or soft computing models, are potentially capable of fitting a nonlinear function or relationships. Identification of model architecture is decisive factor in the simulation and comparison. The identification of model architecture is crucial in ANN model building process. While the input and output of the ANN model is problem dependent, there is no direct precise way to determine the optimal number of hidden nodes (Nayak et al., 2005).The model architecture is selected through a trial and error procedure (Singh et al.. 2004). The fuzzy model, on the other hand, may be considered as a mapping of input space into output space by partitions in the multidimensional feature space in inputs and outputs. Each partition represents a fuzzy set with a membership function.

3.2 Fuzzy rule based system

Fuzzy logic emerged as a more general form of logic that can handle the concept of partial truth. The pioneering work of Zadeh (1965) on fuzzy logic has been used as foundation for fuzzy modeling methodology that allows easier transition between humans and computers for decision making and a better way to handle imprecise and uncertain information. Human being think verbally, not numerically. As the fuzzy logic systems involves verbal statements and, therefore, the fuzzy logic is more in line with human perception (Zadeh, 2000). Fuzzy logic has an advantage over many statistical methods in that the performance of a fuzzy expert system is not dependent on the volume of historical data available. Since these expert systems produce a result based on logical linguistic rules, extreme data points in a small data set do not unduly influence these models. Because of these characteristics, fuzzy logic may be a more suitable method for diffuse pollution forecasting than the usual regression modeling techniques used by many researchers (e,g. Goolsby and Battaglin (1993); Larson and Gilliom (2001); and Tesfamichael et al. (2005) etc.) for estimation of diffuse pollution concentration in streams or other water bodies.

3.2.1 Fuzzy rule based system architecture

The most common way to represent human knowledge is to form it into natural language expression of the type,

$$\text{IF premise (antecedent), THEN conclusions (consequent)} \tag{4}$$

The form in expression (4) is commonly referred to as the IF-THEN rule based form (Ross, 1997). It typically expresses an inference such that if a fact (premise, hypothesis, antecedent) is known, then another fact called a conclusion (consequent) can be inferred or derived. Fuzzy logic systems are rule base systems that implements a nonlinear mapping (Dadone and VanLandingham, 2000) between stresses (represented by consequents) and state variables (represented by antecedents). Creating a fuzzy rule based system may be summarized in four basic steps (Ross 1997; Mahabir et al. 2003; Singh and Singh 2005):

a. For each variable, whether an input variable or a result variable, a set of membership functions must be defined. A membership function defines the degree to which the value of a variable belongs to the group and is usually a linguistic term, such as high or low.
b. Statements, or rules, are defined that relate the membership functions of each variable to the result, normally through a series of IF–THEN statements.
c. The rules are mathematically evaluated and the results are combined. Each rule is evaluated through a process called implication, and the results of all of the rules are combined in a process called aggregation.
d. The resulting function is evaluated as a crisp number through a process called defuzzification.

Subjective decisions are frequently required in fuzzy logic modeling, particularly in defining the membership functions for variables. In cases such as in this study, where large data sets are not available to define every potential occurrence scenario for the fuzzification of model, expert opinion is used to create logic in the rule base system.

3.2.2 Membership functions

Membership functions used to describe linguistic knowledge are the enormously subjective and context dependent part of fuzzy logic modeling (Vadiee, 1993). Each variable must have membership functions, usually represented by linguistic terms, defined for the entire range of possible values. The key idea in fuzzy logic, in fact, is the allowance of partial belongings of any object to different subsets of universal set instead of belonging to a single set completely. Partial belonging to a set can be described numerically by a membership function which assumes values between 0 and 1 inclusive. Intuition, inference, rank ordering, angular fuzzy sets, neural networks, genetic algorithms, and inductive reasoning can be, among many, ways to assign membership values or functions to fuzzy variables (Ross, 1997). Fuzzy membership functions may take on many forms, but in practical applications simple linear functions, such as triangular ones are preferable due to their computational efficiency (Khrisnapuram, R¸1998). In this study, triangular shapes are utilized to represent the membership functions.

3.3 Fuzzy c-means partitioning

Fuzzy rule based models represent the system behaviour by means of if then fuzzy rules. The basic requirement of fuzzy rule based model is to fuzzify or partition the inputs and outputs representation of a physical system. Assigning the number, shape, overlaps etc. of membership functions is most complex part of the fuzzy rule based model building. In most of the cases the optimality of the membership assigned to different fuzzy variables are not guaranteed. FCM is one of the methods to determine the fuzzy partitions of the available data sets into a predetermined number of groups. The data points are divided into group of points that are close to each other. Each data point belongs to a group or cluster with a membership function. Closeness between data points is defined by a metric distance or data center, and each metric yields a different portioning. This cluster centers are utilized in assigning overlaps of triangular shape membership function in this study.

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. The FCM method (developed by Dunn (1973) and improved by Bezdek (1981)) is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C_N} u^m{}_{ij} \left\| x_i - c_j \right\|^2 , 1 \le m < \infty \tag{5}$$

where m is any real number greater than 1, uij is the degree of membership of xi in the cluster j, xi is the ith of d-dimensional measured data, cj is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. The N represents total number of data points, and CN represents the total number of fuzzy centers. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership *uij* and the cluster centers *cj* by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C_N} \left(\frac{\left\| x_i - c_j \right\|}{\left\| x_{i-c_k} \right\|} \right)^{\frac{2}{m-1}}} \tag{6}$$

$$c_j = \frac{\sum_{i=1}^{N} u^m{}_{ij} \cdot x_i}{\sum_{i=1}^{N} u^m{}_{ij}} \tag{7}$$

This iteration will stop when maxij{│uijk+1-uijk │}< ε , where ε is a termination criterion between 0 and 1, where as *k* are the iteration steps. This study used FCM algorithm (Matlab version 6.5), and ε is equal to 0.1 - 10^{-5} to obtain the pre-specified fuzzy centers.

This study implements FCM algorithm (Matlab version 6.5), m=2, and ε equal to 10^{-5} to obtain the pre-specified fuzzy centers.

3.4 Fuzzy rule based system with FCM for estimation of diffuse pollution concentration in streams

The watershed of the streams plays a vital role in influencing the diffuse pollution concentration in the streams. Basic Steps 1 through Steps 4 as discussed earlier in section Rule Based System are implemented by partitioning the input and output spaces into fuzzy regions with FCM, generation of fuzzy rules from available data pairs, assigning a degree to each rule, construction of a combined fuzzy rule base, and mapping from the input space to the output space using the rule base and a defuzzification (Wang and Mendel, 1992).

The vector AC and AL as represented by equation (2) are characterized for the specified watershed of the streams. As explained earlier, AC represents the vector of applied agricultural chemical characteristics such as type of agricultural chemical (insecticide, herbicides etc.), application rate, application season etc. The AL is the land use patterns such as type of crop grown, percentage of cropped area, etc. and C is the stream agricultural diffuse pollution observed concentration measurement values. Patterns were generated using a known set of input-output data pairs. The input data pairs AC and AL values and corresponding output values of C for a particular year constitutes a pattern. While AC and AL are constant for a particular year, the C is temporally and spatially varying at each of the monitoring station sites.

Fuzzy rules are building-blocks of fuzzy rule base systems. Partitioning the fuzzy variables into linguistic variables is necessary step towards designing the rule base system. Fuzzy partitions for the input and output variables are defined or generated according to the type of data as discussed in the membership section (Singh, 2008). In this work, FCM model is utilized to supply optimum number data centers to partition the input and output fuzzy variables.

It is absolutely possible to obtain the redundant and inconsistent rules from the data patterns having same antecedent parts. As mentioned, each rule is assigned a degree or weight by multiplying the membership functions of inputs and outputs for that rule. In the standard approach the rule having largest degree is adopted (Wang and Mendel, 1992). As an improvement, the degree of each rule is multiplied by a redundancy index to obtain the effective degree for that rule. The redundancy index may be defined as:

$$\text{Redundancy Index (R.I.)} = \frac{\mathrm{r_i}}{T_r} \tag{8}$$

where, ri represents the redundant rule with same i antecedents; and Tr represents the sum of all the redundant rules. Final fuzzy rule base includes the rules having the highest effective degree.

The fuzzy inference mechanism uses the fuzzified inputs and rules stored in the rule base for processing the incoming inputs data and produces an output. The fuzzy rules are processed by fuzzy sets operations as discussed in rule based section as basic steps for fuzzy rule base system. The fuzzy rule based design is accepted to be satisfactorily completed when its performance during training and testing satisfies the stopping criteria based on some statistical parameters.

3.5 ANN based methodology for estimation of diffuse pollution concentration in streams

The ANN learns to solve a problem by developing a memory capable of associating a large number of example input patterns, with a resulting set of outputs or effects. ANN is discussed in ASCE Task Committee (2000), etc. An overview of artificial neural networks and neural computing, including details of basics and origins of ANN, biological neuron model etc. can be found in Hassoun (1999), Schalkoff (1997), and Zurada (1997). The details of ANN model building process and selection of best performing ANN model for a given problem is available in (Singh et al., 2004).

As illustrated in the fuzzy model building for estimation of diffuse pollution concentrations in streams, the AC and AL values for a particular year in a watershed are inputs, and corresponding C values in the stream is out put for the ANN model. The values of AC, AL and C for a particular year constitute a data pattern. A standard back propagation algorithm (Rumelhart et al., 1986) with single hidden layer is employed to capture the dynamic and complex relationship between the inputs and outputs utilizing the available patterns. The ANN architecture that perform better than other evaluated architectures based on certain performance evaluation criteria, both in training and testing, was selected as the final architecture.

3.6 Performance evaluation criteria

The performance of the developed models are evaluated based on some performance indices in both training and testing set. Varieties of performance evaluation criteria are available (e.g. Nash and Sutcliffe 1970; WMO 1975; ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models1993 etc.) which could be used for evaluation and inter comparison of different models. Following performance indices are selected in this study based on relevance to the evaluation process. There can be other criteria for evaluation of performance.

3.6.1 Correlation coefficient (R)

The correlation coefficient measures the statistical correlation between the predicted and actual values. It is computed as:

$$R = \frac{\sum_{i=1}^{n}(Xai - \bar{X}ai)(Xpi - \bar{X}pi)}{\sqrt{\sum_{i=1}^{n}(Xai - \bar{X}ai)^2 \sum_{i=1}^{n}(Xpi - \bar{X}pi)^2}} \tag{9}$$

where Xai and Xpi are measured and computed values of diffuse pollution concentration values in streams; $\bar{X}ai$ and $\bar{X}pi$ are average values of Xai and Xpi values respectively; i represents index number and n is the total number of concentration observations.

The correlation coefficient measures the statistical correlation between the predicted and actual values. A higher value of R means a better model, with a 1 meaning perfect statistical correlation and a 0 meaning there is no correlation at all.

3.6.2 Root sean square error (RMSE)

Mean-squared error is the most commonly used measure of success of numeric prediction, and root mean-squared error is the square root of mean-squared-error, take to give it the same dimensions as the predicted values themselves. This method exaggerates the prediction error - the difference between prediction value and actual value of a test case. The root mean squared error (RMSE) is computed as:

$$RMSE = \sqrt{\frac{1}{n}(\sum_{i=1}^{n}(Xai - Xpi)^2)} \tag{10}$$

For a perfect fit, Xai = Xpi and RMSE = 0. So, the RMSE index ranges from 0 to infinity, with 0 corresponding to the ideal.

3.6.3 Standard error of estimates (SEE)

The standard error of estimate (SEE) is an estimate of the mean deviation of the regression from observed data. It is defined as (Allen, 1986):

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(Xai - Xpi)}{(n-2)}} \tag{11}$$

3.6.4 Model efficiency (Nash–Sutcliffe coefficient)

The model efficiency (MENash), an evaluation criterion proposed by Nash and Sutcliffe (1970), is employed to evaluate the performance of each of the developed model. It is defined as:

$$\mathrm{ME_{Nash}} = 1.0 - \frac{\sum_{i=1}^{n}(Xa_i - X_{pi})^2}{\sum_{i=1}^{n}(X_{ai} - \bar{X}ai)^2} \tag{12}$$

A value of 90% and above indicates very satisfactory performance, a value in the range of 80–90% indicates fairly good performance, and a value below 80% indicates an unsatisfactory fit.

4. Data synthesis and architecture identification of models

In this work, the diffuse pollution concentration in stream is considered due to herbicide atrazine application in corn fields of the watershed. Concentration measurements data were obtained from the National Water Quality Assessment (NAWQA) program of the U S Geological Survey (USGS) (http://water.usgs.gov/nawqa/naqamap.html) for the period 1992 to 2002. The stream considered is White River, and monitoring site for the atrazine concentration measurement, is Hazeltone (Crawford, C.G, 1995), the outlet site of the watershed of White River Basin in Indiana State. At Hazeltone site, Latitude is 38°29'23", and Longitude is 87°33'00" and Drainage area 11,305.00 square miles. The White River basin is a part of the Mississippi River system where the application of atrazine accounts for 24 percent of all agricultural herbicides. The major agricultural chemical characteristics, AC, which contribute to the atrazine concentration at the watershed outlet are identified as its application rate (lb/Acre) and application time. The major land use patterns, AL, is the extent of cropped area (percentage of cultivated area (Pareira, 1990; Crawfard, 2001; and Capel and Larson, 2001).

Time series of data (average monthly values) from 1992-2001 are utilized for model building and validation. The major agricultural chemical characteristics, AC, which contribute to the atrazine concentration at the watershed outlet are identified as its application rate (lb/acre) and application time. The major land use pattern, AL, is the extent of cropped area (percentage of cultivated area (Crawford, 2001, 1995).These data are utilized for identification of fuzzy and ANN based models architectures by applications of the methodologies discussed in previous sections. The performance evaluations criteria are utilized to judge the predictive capability of the best performing fuzzy and ANN models. The procedure of developing fuzzy logic rule based model is implemented using the data of atrazine application rate as first input, atrazine application season as second input, and the percentage area applied with atrazine as third input. The atrazine concentration measurement values observed at the monitoring site is the output for the fuzzy rule based model. The weighted average of herbicide application rates and percentage of area applied of the corn and soybean cropped area are given in Table 1. The seven years data (1992-1998) are utilized for training and the three years data (1999-2001) (Table 1) are utilized for testing models.

Year	Weighted Percentage Area	Application Rate (lb/Acre)
1992	79	1.35
1993	91	1.31
1994	87	1.35
1995	87	1.31
1996	91	1.31
1997	84	1.33
1998	89	1.36
1999	91	1.26
2000	80	1.41
2001	94	1.35

Table 1. Agricultural Herbicide Atrazine Application Rate and Percentage Area Applied for the Corn Crop.

4.1 Evaluation of fuzzy c-means centers

The FCM model represented by equation (5) is used to partition the input data into fuzzy partitions. The FCM algorithm is implemented using MATLAB version 6.5 for ε equal to 10^{-5} to obtain the pre-specified fuzzy centers. The 3, 4, and 5 fuzzy centers for the inputs application rate and weighted percentage area obtained using the FCM model is shown in Table 2. Instead of iterating for the optimal number of fuzzy centers, a prior knowledge about the fuzzy partitioning for the fuzzy rule based models were utilized in implementing fuzzy c-means algorithm.

Fuzzy Partition centers by FCM Model		
Fuzzy Partitions	Input Application Rate (lb/Acre)	Application Rate (lb/Acre)
3-Fuzzy Centers	1.26 1.31 1.37	80.38 86.68 90.75
4-Fuzzy Centers	1.26 1.31 1.33 1.36	79.50 84.02 87.21 90.88
5-Fuzzy Centers	1.26 1.31 1.33 1.35 1.41	80.00 86.67 87.00 89.17 91.0

Table 2. Different Fuzzy Partition Centers Using FCM Model

4.2 Training and testing the fuzzy rule based model with FCM

The seven years data (1992-1998) are utilized for training and the three years data (1999-2001) are utilized for testing the fuzzy rule based model with FCM. The model is assumed to be performing satisfactory when model efficiency coefficient (MENash) as given by equation (12) is greater than 90 percent, and other performance indices are also improved. Although arbitrary, it may be used as stopping criteria to limit the processing of large number of rules with increase in linguistic fuzzy variables for the inputs.

Performance of fuzzification of inputs application rate and weighted percentage area were studied by assigning 3, 5, and 7 fuzzy variables without using FCM (Singh, 2008). Though performance of fuzzifiction with 7 variables worked better than fuzzification with 3 and 5 variables; fuzzification by 5 fuzzy variables are comparable to fuzzification with 7 variables as shown in Table 3. Fuzzy rule based models with 3, 5 and 7 fuzzy variables are represented by Fuzzy_3M, Fuzzy_5M, and Fuzzy_7M models respectively in the Table 3. As 3 partitions are not adequate, four fuzzy partitions were specified for the use of fuzzy rule based system with FCM model. The four centers as shown in Table 2, obtained using FCM are partitioned into four linguistic fuzzy variables as low, medium, high, and very high. A sample schematic representation of membership function is shown for the input atrazine application rate in Figure 1.

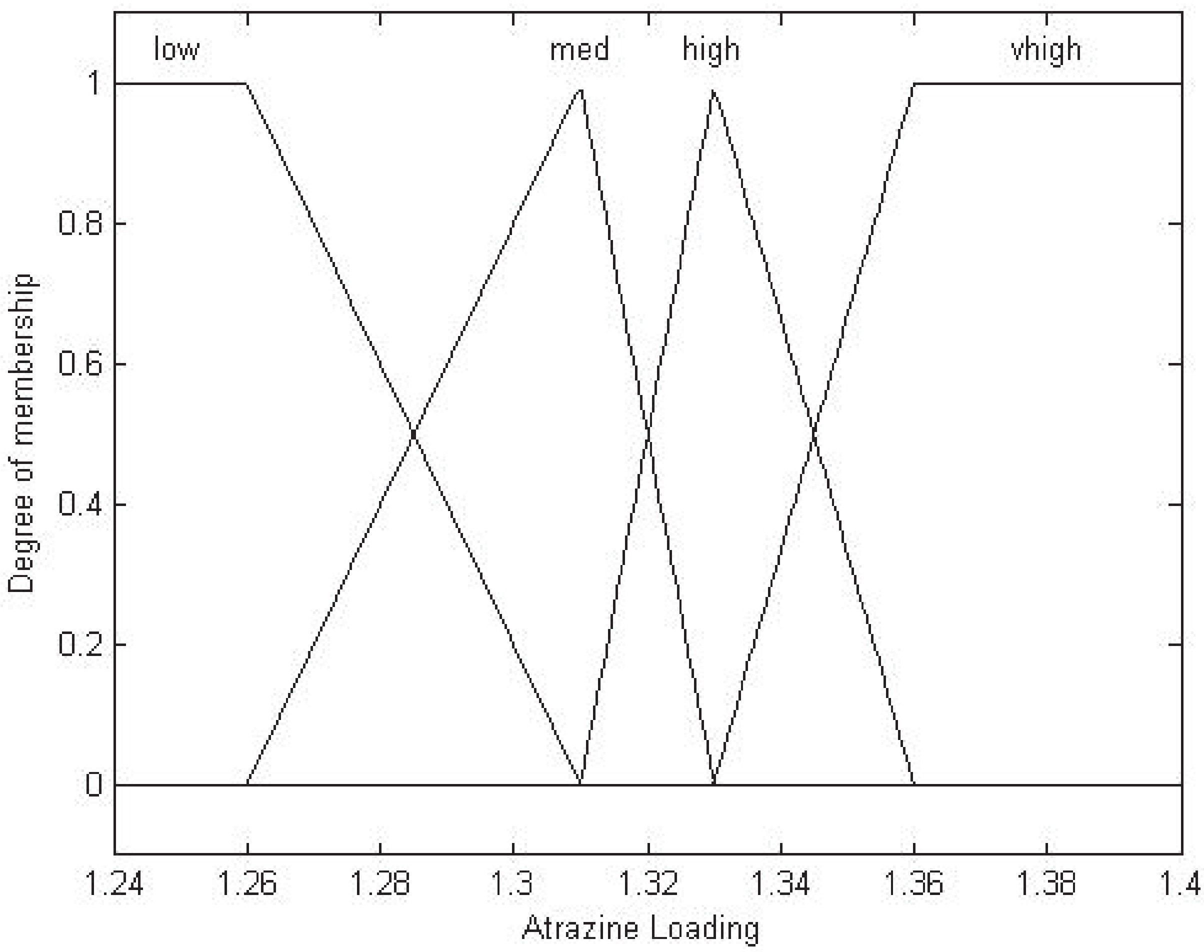


Fig. 1. A sample representation of linguistic variables membership function for first input.

The input application season is assigned 12 fuzzy variables, S1-S12 corresponding to each month of a year. The output concentration measurement values of atrazine is represented by 25 fuzzy centers by FCM model and represented by fuzzy variables, C1-C25, so that all the ranges of atrazine concentration measurement values in the data set for the period 1992-2001, is adequately represented. All the fuzzy variables in inputs and outputs are represented by triangular shape, except at the domain edges, where they are semi trapezoidal. This representation has been selected based on literature due to their computational efficiency (Khrisnapuram R 1998; Guillaume and Charnomordic, 2004). A sample representation of the membership functions is shown in Figure 1 for the first input. Of course, other divisions of the inputs and output domain regions and other shapes of membership functions are possible. The total number of rules in case of 4 linguistic variables for inputs application rate and weighted percentage area, and 12 fuzzy variables for seasons are 192. The total number of rules was much high i.e. 588 when 7 fuzzy variables were used for inputs application rate and weighted percentage area. The model building process is completed by creating combined fuzzy rule base using inputs-output pair values of training set data. Finally, the defuzzification converts fuzzy output produced by the fuzzy rule base model as crisp output corresponding to any new inputs.

5. Concentration measurement estimation results

The performance of the FCM based fuzzy rule based model is evaluated based on performance indices as described in performance evaluation criteria. These include root mean square error (RMSE), correlation coefficient (R) between the actual and estimated monthly average concentration measurement values of atrazine herbicides, standard error of estimate (SEE) and MENash. The performance evaluation results of the fuzzy rule based model with four fuzzy variables obtained using FCM, represented as Fuzzy_4_FCM, is also compared with that of the fuzzy rule based models with 3, 5, 7 linguistic variables for both of the input 1 and input 3. The performance of the Fuzzy_4_FCM model is also compared with solution results of an artificial neural network (ANN) based model using back propagation algorithm (Rumelhart et al. 1986) as represented by ANN_M in Table 3.

Models	Training Error (1992-198)				Testing Error (1999-2001)			
	RMSE	**R**	**SSE**	**ME_{Nash}**	**RMSE**	**R**	**SSE**	**ME_{Nash}**
Fuzzy_3M	1.318	0.891	1.377	0.550	0.703	0.886	0.771	0.623
Fuzzy_5M	0.836	0.969	0.837	0.894	0.455	0.952	0.498	0.855
Fuzzy_7M	0.706	0.970	0.775	0.915	0.342	0.975	0.375	0.914
ANN_M	1.153	0.918	1.264	0.752	0.906	0.759	0.993	0.446
Fuzzy_4M_FCM	0.492	0.998	0.539	0.967	0.725	0.968	0.416	0.901

Table 3. Comparison of training and testing errors for different models.

It can be noted from the Table 3 that the error statistics are better for Fuzzy_4M_FCM model than those of Fuzzy_3M, Fuzzy_5M and ANN_M model in both the training and testing in prediction in atrazine concentration measurement values. Its performance is even better than Fuzzy_7M model in training. Model efficiency (MENash) in training is 94.3 percent whereas it is 91.5 percent for Fuzzy_7M model. Similarly, RMSE, R, and SSE values are also comparable. In testing, results are also comparable though error statistics for Fuzzy_7M model is slightly better than Fuzzy_4_FCM. Thus, the FCM optimized fuzzy membership functions partitions in Fuzzy_4_FCM model are performing comparable to almost double the fuzzy partitions without FCM in Fuzzy_7M model. Figure 2 shows better RMSE value by Fuzzy_4_FCM model in comparison to other models.

It can also be noted from Table 3 that performances of fuzzy rule based model is better than those obtained using an ANN model with 2 inputs (atrazine application rate and weighted percentage area), 12 outputs (average monthly concentration measurements), and 11 hidden nodes (selected on the basis of experimentation) represented by ANN_M model. The poor performance by ANN_M model may be due to inadequate training patterns for experimentation, as the total number of free parameters become more than the number of training patterns even for 1 hidden node in hidden layer.

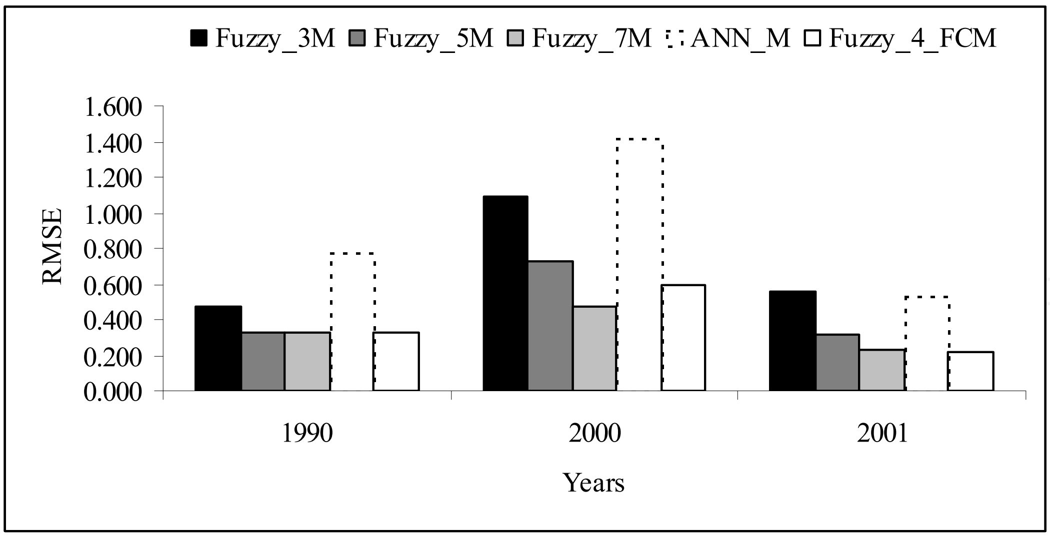


Fig. 2. Performance comparison of models.

Scatter plots of average monthly observed and predicted atrazine concentration measurement in the stream for model Fuzzy_4_FCM are plotted for the testing period 1999, 2000, and 2001. Comparison of actual and model estimated values are also presented for average monthly variations of atrazine concentration in the stream during the testing period, 1999-2001. Figure 3 represents scatter plot, and Figure 4 represents comparison of actual and Fuzzy_4_Model estimated values for the period 1999. Scatter plots between the observed and Fuzzy_4_FCM predicted average atrazine concentration measurement values in stream followed a 1:1 line except for a few cases of high magnitudes. The high values of coefficient of determination, R^2 (0.933), indicate that there is a good match between the observed and model predicted atrazine concentration. Figure 4 shows a comparison of observed and, Fuzzy_4_FCM model predicted average monthly atrazine concentration measurement values in the stream. The observed and Fuzzy_4_FCM predicted values match well except for the occurrence of peak value.

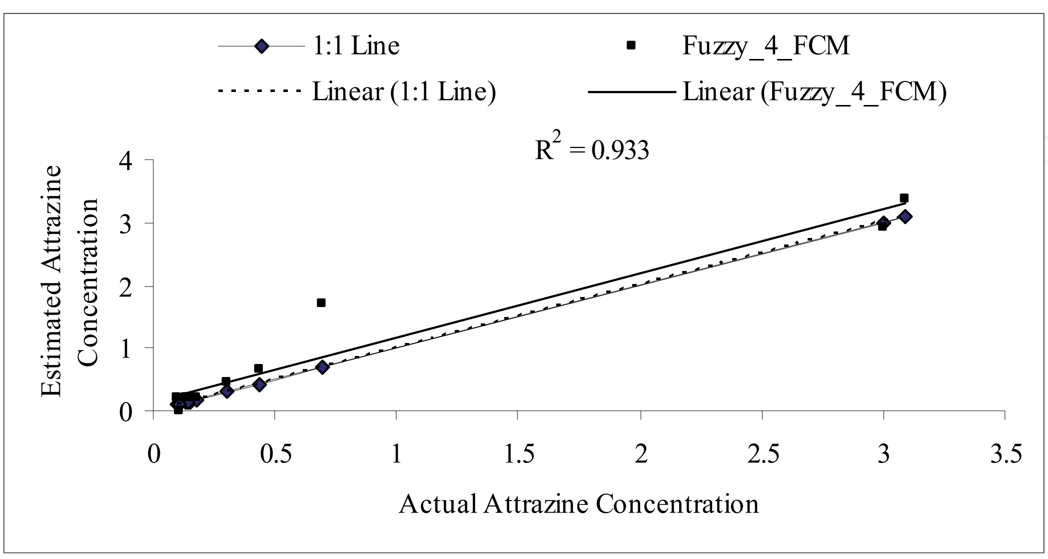


Fig. 3. Scatter plot of observed and Fuzzy_4_FCM Model predicted average monthly atrazine concentration for the testing period year 1999.

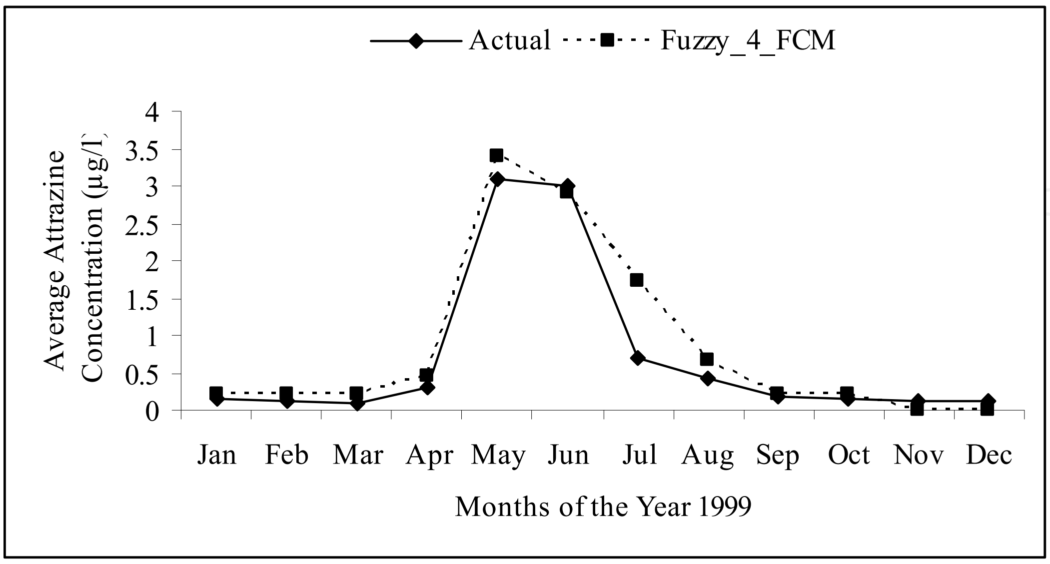


Fig. 4. Comparison of observed and Fuzzy_4_FCM predicted average monthly atrazine concentration for the testing period year 1999.

Figure 5 represents scatter plot of observed and Fuzzy_4_FCM predicted values, and Figure 6 represents comparison of observed and Fuzzy_4_FCM predicted atrazine concentration values for the period 2000. Scatter plots between the observed and Fuzzy_4_FCM predicted average atrazine concentration measurement values in stream followed a 1:1 line with R^2 value of 0.95. In this case though initial and final months values matches well, intermediate months values including peak value does not mach well as shown in Figure 6.

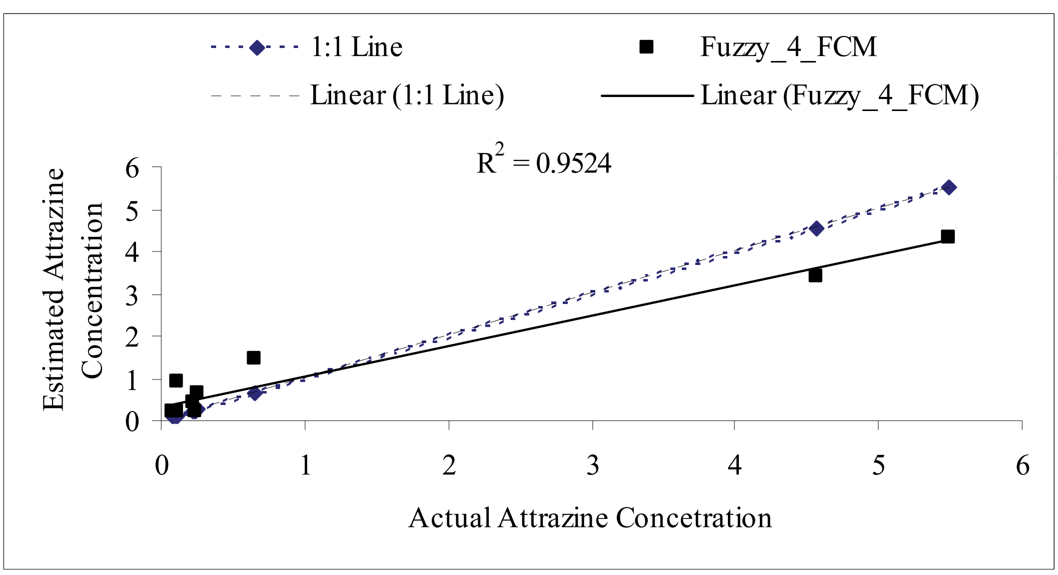


Fig. 5. Scatter plot of observed and Fuzzy_4_FCM Model predicted average monthly atrazine concentration for the testing period year 2000.

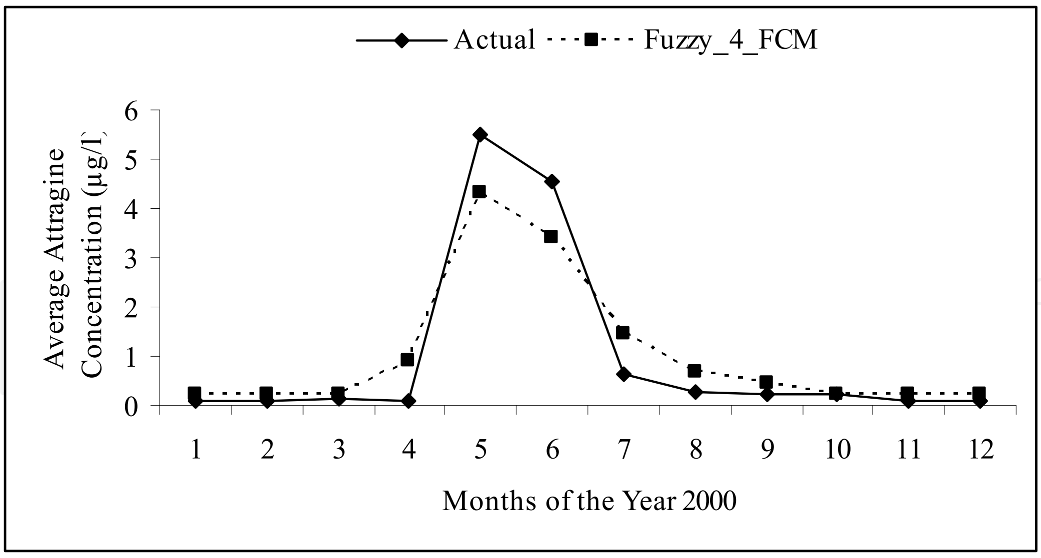


Fig. 6. Comparison of observed and Fuzzy_4_FCM predicted average monthly atrazine concentration for the testing period year 2000.

Figure 7 represents scatter plot of observed and Fuzzy_4_FCM predicted values, and Figure 8 represents comparison of observed and Fuzzy_4_FCM model predicted atrazine concentration values for the period 2001. Scatter plots between the observed and Fuzzy_4_FCM predicted average atrazine concentration measurement values in stream followed a 1:1 line with high value R^2 (0.93).

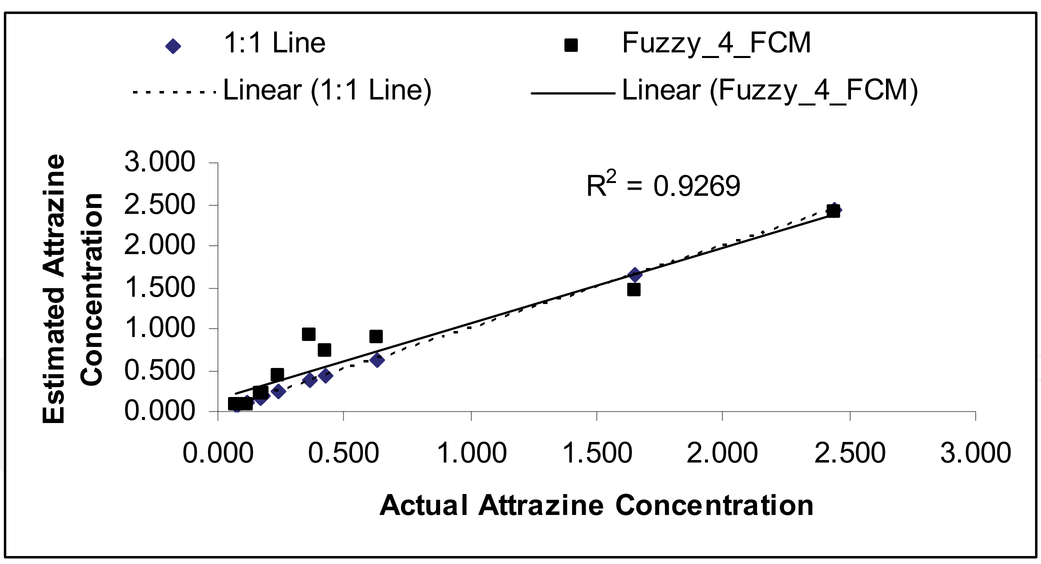


Fig. 7. Scatter plot of observed and Fuzzy_4_FCM Model predicted average monthly atrazine concentration for the testing period year 2001.

6. Discussion of results

The performance evaluation results presented in this study establish the potential applicability of the developed methodology in estimation of monthly atrazine concentration measurement values using fuzzy rule based models with FCM. However, the comparative

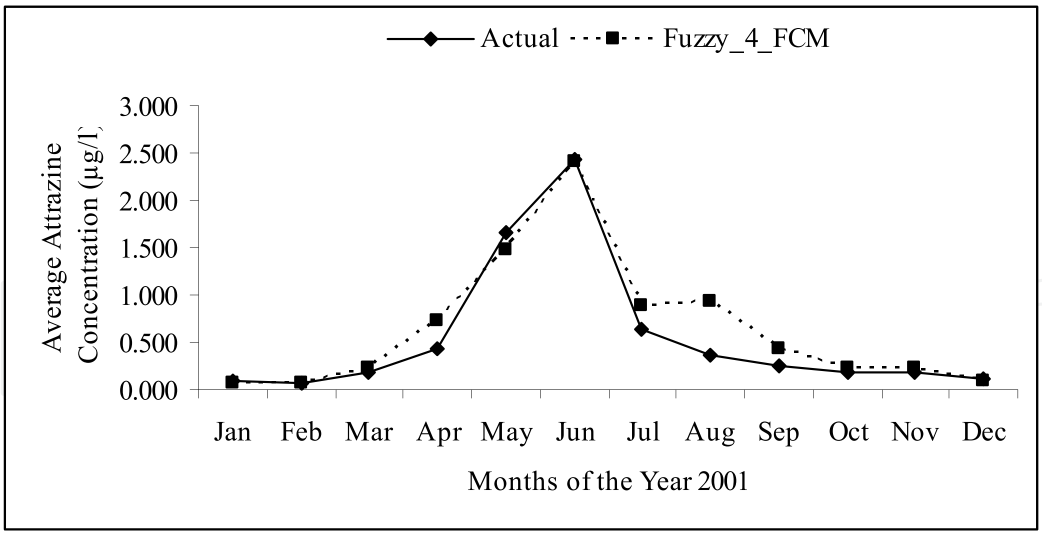


Fig. 8. Comparison of observed and Fuzzy_4_FCM model predicted average monthly atrazine concentration for the testing period year 2001.

performance of the methodology in different evaluation periods, under or over prediction of peak values, fuzzy rule based model control parameters (shape, total number of fuzzy centers, overlaps etc. of membership functions; fuzzy set operations i.e, defuzzification methods etc.) needs to be investigated further.

The performance of fuzzy rule based model with FCM is better than those without FCM model with even more number of fuzzy partitions. This is inferred by comparison of performances of Fuzzy_4_FCM model with Fuzzy_3M, Fuzzy_5M, and Fuzzy_7M models. In all the evaluation results obtained by Fuzzy_4_FCM model for the period 1999-2001, the R^2 values from scatter plots, and MENash values obtained from observed and model predicted values are high (around 0.9). This implies good match between the observed and model predicted values. The fuzzy rule with FCM model also performed better than the ANN based model. It establishes that the developed fuzzy rule based model with FCM is potentially suitable for estimation of concentration measurement values with limited data availability. The performances of the developed models are better in comparison to performance of regression models developed for the Mississippi River Systems (Battaglin and Goolsby, 1997). Their study show that multiple linear regression models estimate the concentration of selected agricultural chemicals with maximum R-squared value is 0.514, and in the case of atrazine, R-squared value is 0.312. In this study, almost all the developed models have R-squared value greater than 0.55. However, this comparison is limited as the White River basin considered in this study is only a part of (one of 10 basins) of Mississippi River Systems considered by them (Battaglin and Goolsby, 1997).

The estimation results obtained using fuzzy rule based models are encouraging but not conclusive. In almost all the evaluations, though initial months and final months concentration measurement values matches well, the intermediate values including the peak values are either over predicted or under predicted except for the year 2001 where peak predicted value matched well with the observed value. As the intermediate months, from April to July observes most of the changes in atrazine observed concentration measurement values, the same dynamics are exactly not reflected in model predictions. Thus, though the FCM model works better than ANN model in case of limited data availability, its performance is also affected due to limited data sets. In the present study, the inputs were assigned with triangular shape. Further improvement in the performance of the methodology may be possible with more extensive evaluations of membership functions shape, number of data centers for membership functions for each variables, and overlap between two membership functions. Present methodology utilized centroid method for defuzzification. Performance of other defuzzification method also need to be investigated. The error in prediction of peak values shows the limitation of the methodology. However, these results show potential applicability of the proposed methodology. The main advantage of the developed methodology is incorporate some prior knowledge into the model frame work, and its ability to perform in case of limited availability of data than other methods such as ANN.

7. Conclusions

The present study describes the framework for evaluating average monthly concentration of agricultural non point source pollution due to herbicide atrazine in streams by fuzzy rule based model with FCM utilizing limited amount of data. The values of statistical performance evaluation criteria indicate the model is able to simulate the behaviour of diffuse pollution sources from agricultural fields like attrazin in streams. The fuzzy rule based model with FCM performs comparatively better than the fuzzy rule based model without FCM and even with more fuzzy partitions. The proposed methodology also performs better than the ANN model when applied to the same problem. However, the model predicts with lesser accuracy for the intermediate months concentration measurement values including peak values. An extensive evaluation of the effect of more number of FCM based fuzzy centers and shapes of membership functions may fully establish the applicability of the methodology.

However, the proposed fuzzy rule based approach with FCM uses least amount of information in terms of number of inputs required, incorporate prior knowledge about fuzzy partitions, and also uses linguistic variables which make it relatively easy to interpret the rules. Prior knowledge about the physical system in the form of rule base can also be directly incorporated in the suggested approach. This preliminary study shows that the developed fuzzy rule based approach with FCM is potential suited to estimation of diffuse pollution concentration like atrazine in streams.

8. References

ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models (1993) Criteria for evaluation of watershed models. *J. Irrig. Drain. Eng.* ASCE 119(3): 429–442.

Arnold, J.G., Allen, P.M., Bernhardt, G. (1983) A comprehensive surfacegroundwater flow model. *Journal of Hydrology* 142: 47–69.

Baker D.B., Richards R.P. (1990) Transport of soluble pesticides through drainage networks in large agricultural river basins, In: Kurtz D.A. (Ed.), Long Range Transport of Pesticides. Lewis Publishers, Inc., Chelsea, MI.

Battaglin, W. A. and Goolsby, D. A., 1997. Statistical modeling of agricultural chemical occurrence in Midwestern Rivers. *Journal of Hydrology*, 196, 1-25.

Bezdek J. C. (1981) Pattern Recognition with Fuzzy Objective Function Algoritms. Plenum Press, New York.

Capel Paul D., Larson Steven J. (2001) Effect of Scale on the Behavior of Atrazine in Surface Waters. *Envron. Sci. Technol.* 35(4): 648-657, 2001.

Charbeneau R., Barrett M. (1998) Evaluation of methods for estimating storm water pollutant loads, *Water Environment Research:* 70(7), 1295-1302.

Crawfard C.G. (2001) Factors Affecting Pesticide Occurrence And Transport In A Large Mid-Western River Basin. *Journal of American Water Res. Asstn.* 37(1): 1-15, 2001.

Crawford, C.G. (1995) Occurrence of pesticides in the White River, Indiana, 1991-95: U.*S. Geological Survey Fact Sheet No. 233-95.*

Coupe R. H., Goolsby D. A., Iverson J. L., Zaugg S.D., Markovchick D. J. (1995) Pesticide, Nutrient, Streamflow and Physical Property Data for the Mississippi River, and Major Tributaries, April 1991- September, 1992. *U.S. Geol. Surv. Open-File* Rep. 93-657.

Dadone P, VanLandingham H.F. (2000) On the differentiability of fuzzy logic systems. *Proceedings of IEEE Conference on Systems, Man and Cybernetics*, Nashville, TN, 2703-2708.

D'Arcy B.J., Ellis J.B., Ferrier R.C., Jenkins A., Dils R. (eds.) (2000) Diffuse pollution impacts, the environmental and economic impacts of diffuse pollution in the UK, *Chartered Institution of Environmental Management(CIWEM), Lavenham Press.*

Dunn J. C. (1973) A Fuzzy Relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3: 32-57

Goolsby D. A., Battaglin, W. A. (1993) Occurrence, distribution, and transport of agricultural chemicals in surface waters of the midwestern United States, In: Goolsby D. A., Boyer L.L, Mallard G.E. (Eds.), Selected papers on agricultural chemicals in Water Resources Midcontinental United States. *U S Geological Survey Open*-File Rep. 93-418: 1-24

Goolsby D. A., Battaglin, W. A. (1995) Occurrence, and distribution of pesticides in rivers of the midwestern United States, In: Leng, M.L., Leovey, E.M.K., Zubkoff, P.L.(Eds.), *Agrochemical Environmental Fate: State of the Art.* CRC Press Inc., Boca Raton, FL: 159-173.

Guillaume S, Charnomordic B. (2004) Generating an interpretable family of fuzzy partitions from data. *IEEE transactions on Fuzzy Systems,* 12 (3): 324-335.

Jha, R., Ojha, C.S.P, Bhatia, K.K.S., 2005. Estimating nutrient outflow from agricultural watersheds to the river Kali in India. *Journal of Environmental Engineering*, ASCE, 131(12), 1706–1715.

Khrisnapuram R. (1998) Membership function elicitation and learning. In: Ruspini E.H., Bonissone P.P., Pedrryez W. (eds.). *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Dirac House, Temple Bath, Bristol.

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*. 3(5): 1349-1367.

Mahabir C, Hicks, F.E., Fayek A.R. (2003) Application of fuzzy logic to forecast seasonal runoff. *Hydrol. Process.* 17: 3749-3762.

Mahar, P.S., and Datta, B. (2000) Identification of pollution sources in transient groundwater system. *Water Resource Management* 14(6): 209 - 227.

Nash J. E., and Sutcliffe, J. V. (1970) River flow forecasting through conceptual models. Part 1-A: Discussion principles. *J. Hydrol.* 10: 282–290

Novotny V., Water Quality (2nd edition) (2003), *Diffuse Pollution & Watershed Management.* John Wiley and Sons, New York. ISBN 0-471-39633-8.

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*. 24 (9): 1400-1406.

Ross T.J. (1997) *Fuzzy Logic with Engineering Applications*. McGraw-Hill, Book Co., Singapore.

Rumelhart, D.E., Hinton, G.E., Williams, R.J. (1986) Learning internal representation by error propagation. *Parallel Distributed Processin. 1:* 318-362, MIT Press, Cambridge, Mass.

Schottler S. P., Eisenreich S. J., Capel P. D. (1994) Atrazine, Alachlor, and Cyanazine in a Large Agricultural River System. *Environ. Sci. Technol.,* 28, p 1079-1089.

Scribner, E. A., Goolsby D. A., Thurman E. M., Meyer M. T., Pomes M. L. (1994) Concentrations of Selected Herbicides, Two Triazine Metabolites, and Nutrients in Storm Runoff From Nine Stream Basins in the Midwestern United States, 1990-92. *U.S. Geol. Surv. Open-File Rep.* 94-396.

Singh, R. M. (2008). Fuzzy Rule Based Estimation of Agricultural Diffuse Pollution Concentration in Streams. *Journal of Environmental Science and Engineering* (ISSN036-827 X), 50 (2), 147-152.

Singh, R. M., Datta, B. (2006a) Identification of unknown groundwater pollution sources using genetic algorithm based linked simulation optimization approach. Journal of Hydrologic Engineering ASCE 11(2): 101-109.

Singh R. M., and Datta, B. (2006b) Artificial neural network modeling for identification of unknown pollution sources in groundwater with partially missing concentration observation data. *Water Resources Management (Springer)*, 21 (3), 557-572.

Singh, R.M., Singh, R. (2005). Agricultural diffuse pollution estimation using fuzzy rule based model. In: *Proceedings of the International Symposium on Recent Advances in Water Resources Development and Management (RAWRDM-2005),* IIT Roorkee, India, 1010-1018.

Smith, R.M. S., Wheater H.S. (2004) Multiple objective evaluation of a simple phosphorus transfer model *Hydrol. Process.* 18(9): 1703-1720.

Tesfamichael A.A., Caplan A.J., Kaluarachchi J.J. (2005) Risk-cost-benefit analysis of atrazine in drinking water from agricultural activities and policy implications. *Water Resources Research,* 41, W05015, 1-13.

U.S. Environmental Protection Agency (USEPA). 2001. Drinking water exposure assessment for atrazine and various chloro-triazine and hydroxyl-triazine degradates, report, Environ. Fate and Effects Div., Washington, D.C.

U.S. Environmental Protection Agency (USEPA). 2002. Cost of illness handbook, report, Off. of Pollut. Prevention and Toxics, Washington, D.C.

Vadiee, N. (1993). Fuzzy rule-based expert systems –I and II. In M. Jamshidi, N. Vadiee and T. Ross (eds.), *Fuzzy logic and control: software and hardware applications*, Prentice Hall, Englewood Cliffs, N.J., Chapters 4 and 5.

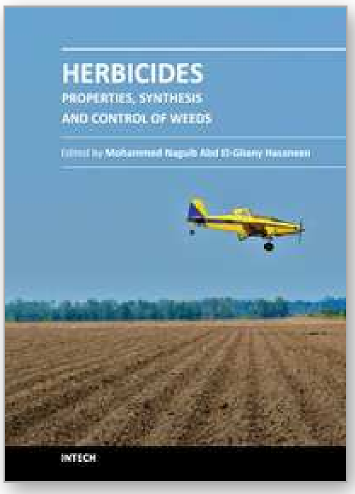
Wang Li-Xin, and Mendel J.M. (1992) Generating Fuzzy rules by Learning from Examples. *IEEE transactions on Systems, Man and Cybernetics* 22 (6): 1414-1427.

World Meteorological Organization (WMO). (1975) Intercomparison of conceptual models used in operational hydrological forecasting. Tech. Rep. No. 429, WMO, Geneva, Switzerland.

Young, R.A., Onstad, C.A., Boesch, D.D., Anderson, W.P. (1989) AGNPS—a nonpoint-source pollution model for evaluating agricultural watersheds. *Journal of Soil and Water Conservation* 44 (2): 168–173.

Zadeh, L.A. (2000). Towards a Perception-Based Theory of Probabilistic Reasoning- Abstract of a plenary address at the conference of North American Fuzzy Information Processing Society, 2000.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control* 8: 338-353.



Herbicides - Properties, Synthesis and Control of Weeds
Edited by Dr. Mohammed Nagib Hasaneen

ISBN 978-953-307-803-8
Hard cover, 492 pages
Publisher InTech
Published online 13, January, 2012
Published in print edition January, 2012

This book is divided into two sections namely: synthesis and properties of herbicides and herbicidal control of weeds. Chapters 1 to 11 deal with the study of different synthetic pathways of certain herbicides and the physical and chemical properties of other synthesized herbicides. The other 14 chapters (12-25) discussed the different methods by which each herbicide controls specific weed population. The overall purpose of the book, is to show properties and characterization of herbicides, the physical and chemical properties of selected types of herbicides, and the influence of certain herbicides on soil physical and chemical properties on microflora. In addition, an evaluation of the degree of contamination of either soils and/or crops by herbicides is discussed alongside an investigation into the performance and photochemistry of herbicides and the fate of excess herbicides in soils and field crops.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821