1. Introduction: It is, by now, well recognized that Agriculture is a “Soft science”, unlike Physics or Chemistry, which are “Hard sciences”. In the former, there is always some amount of “impreciseness” or “vagueness” or “fuzziness” in the underlying phenomenon, and/or explanatory variables, and/or response variable(s). Therefore, for a more realistic modelling, there is a need to incorporate this aspect in traditional models, like Multiple linear regression model.

2. Fuzzy boom (1987 - Present): Yen and Langari (2004) have thoroughly described the “Fuzzy Boom” since 1987. In 1988, Two large-scale National research projects on “Fuzzy logic” were established by two agencies in Japan, viz. Ministry of International Trade and Industry (MITI) and Science and Technology Agency (STA). The project established by MITI was a consortium called the Laboratory for International Fuzzy Engineering Research (LIFE), which involved 50 companies with a six-year total budget of $5,000,000,000. Panasonic was the first to apply fuzzy logic to a consumer product, a shower head that controlled water temperature. In 1990, Panasonic named their newly developed “Fuzzy controlled” automatic washing machine. Many other home electronic companies followed Panasonic’s approach and introduced Fuzzy vacuum cleaners, Fuzzy rice cookers, Fuzzy refrigerators, and others. The first financial trading system using Fuzzy logic was Yamaichi Fuzzy Fund. It handles 65 industries and a majority of the stocks listed on Nikkei Dow and consists of approximately 800 Fuzzy rules. Another important milestone is the development of first VLSI chip for performing Fuzzy logic inferences. The Fuzzy Logic Toolbox was introduced in 1994 as an add-on component to MATLAB Software Package.

3. Fuzzy linear regression methodology: In conventional regression analysis, deviations between observed and estimated values are assumed to be due to random errors. However, quite often these are due to indefiniteness of structure of a system or imprecise observations. Thus, uncertainty in this type of regression model becomes "fuzziness” and not randomness. Studies dealing with Fuzzy linear regression (FLR) model can be broadly classified into two approaches, viz. (i) Linear programming (LP)-based methods, and (ii) Fuzzy least squares (FLS) methods. In the former approach, proposed by Tanaka et al. (1982), parameters of FLR model:

\[ Y = A_0 + A_1X_1 + \ldots + A_pX_p \]  

where \( A_i = (a_{ic}, a_{iw}) \), \( Y = (y_c, y_w) \) are estimated by minimizing “Total vagueness” of model-data combination, subject to constraints that each data point must lie within estimated value of response variable. This can be visualized as a LP problem and solved by using “Simplex procedure” as discussed below.
Let us consider two sets $X$ and $Y$ and a function $f(x, a)$ which is a mapping from $X$ into $Y$. If parameters are given by fuzzy sets $A$, the function is called a fuzzy function, denoted by $f(x, A)$ mapped from the fuzzy set $A$ can be defined as follows.

**Definition 1:** The fuzzy function is denoted by

$$f: X \rightarrow \mathcal{F}(y); Y = f(x, A)$$

where $\mathcal{F}(y)$ is the set of all fuzzy subsets on $Y$. The fuzzy set $Y$ is defined by the membership function

$$\mu_Y(y) = \begin{cases} \max \{\mu_A(a) \mid (a \mid y = f(x, a)) \neq \emptyset\}, & \text{if } a \neq \emptyset \\ 0, & \text{otherwise} \end{cases}$$

where $A$ is fuzzy set on the space of parameters whose membership function is defined by $\mu_A(a)$.

**Definition 2:** The above fuzzy parameters are defined by

$$\mu_A(a) = \min \{\mu_A(a_j) \mid a = \bigcup_j a_j\}$$

where

$$\mu_A(a_j) = \begin{cases} \left[1 - \frac{\alpha_j - a_j}{c_j}\right], & a_j - c_j \leq a_j \leq a_j + c_j, \quad c_j > 0. \\ 0, & \text{otherwise} \end{cases}$$

Instead of considering the deviation between observation and estimated value $Y_i = \alpha_1 x_1 + \ldots + \alpha_n x_n$ as the observation error, we assume that the deviations depend on the fuzziness of the system structure.

To formulate a fuzzy linear regression model, the following are assumed to hold:

1. The data can be expressed by a fuzzy linear model:

$$Y_i^* = A^* x_i^* = \sum_{j=1}^{n} a_j x_j^*$$

Given $x_i$, $Y_i^*$ can be obtained using the proposition in Tanaka et al. (1982), as a fuzzy set given by

$$\mu_{Y_i^*}(y) = 1 - \frac{|y - x_i^*|}{c|x_i^*|}.$$

2. The degree of the fitting of the estimated fuzzy linear model $Y_i^* = A^* x_i$ to the given data $Y_i = (y_i, e_i)$ is measured by the following index $\hat{h}$ which maximizes $h$ subject to $Y_i^{h} \subset Y_i^{*h}$, where

$$Y_i^{h} = \{y \mid \mu_{Y_i}(y) \geq h\}$$

$$Y_i^{*h} = \{y \mid \mu_{Y_i^*}(y) \geq h\},$$

which are $h$-level sets.

The degree of fitting of the fuzzy linear model to all data $Y_1, \ldots, Y_N$ is defined by $\min_j [\hat{h}]$.

3. The vagueness of the fuzzy linear model is defined by

$$J = c_1 + \ldots + c_n.$$
The problem is explained as obtaining fuzzy parameters $\bar{A}_i^*$ which minimize $J$ subject to $\bar{H}_i \geq H$ for all $i$, where $H$ is chosen as the degree of the fitting of the fuzzy linear model by the decision-maker.

Specifically our problem is to find fuzzy parameters $A = (\alpha_i, c_i)$ which are the solution of the following linear programming problem:

$$\min_{\alpha, c} J = c_1 + \ldots + c_n$$ subject to $c_j \geq 0$

and

$$\alpha^* x_i + (1 - H) \sum_j c_j |x_{ij}| \geq y_i + (1 - H) e_i$$

$$- \alpha^* x_i + (1 - H) \sum_j c_j |x_{ij}| \geq - y_i + (1 - H) e_i$$

for $i = 1, 2, \ldots, N$.

Several software packages, like SAS, LP88, and LINDO are available for solving LP problem numerically by Simplex procedure. However, any standard spreadsheet package, like Microsoft Excel may also be used to solve it manually.

Kandala and Prajneshu (2003) demonstrated applicability of above methodology when the two explanatory variables (viz. Plant height and Leaf area index) and response variable (Dry-matter accumulation) are all crisp but underlying phenomenon is assumed to be fuzzy in nature. It was shown that widths of prediction intervals in respect of Fuzzy linear regression model were much less than those for Multiple linear regression model. Kandala and Prajneshu (2002) had earlier obtained similar results in the situation when the two explanatory variables, viz. Normalized difference vegetation index (NDVI) and Ratio vegetation index (RVI) are highly correlated.

Further, for determining age-length relationship in a fish species, response variable (length) generally lies in an interval for different fish of same age. Kandala and Prajneshu (2004) applied FLR methodology for fitting fuzzy von Bertalanffy growth model with a view to determining age-length relationship in pearl oyster. It may be pointed out that traditional statistical methods are not capable of handling such a situation in which response variable is in intervals. The only way out there is to get rid of interval values for response variable to crisp values either by taking mean or mode, thereby losing a lot of vital information about spread. However, a criticism of Tanaka’s approach is that it is not based on sound statistical principles. Another drawback, as pointed out by Chang and Ayyub (2001) and D’Urso (2003), is that as the number of data points increases, the number of constraints in LP increases proportionately, thereby resulting in computational difficulties.

The second approach based on Fuzzy least squares (FLS) method, was pioneered by Diamond (1988), which as its name suggests, is a fuzzy extension of Least squares method based on a new defined distance on the space of fuzzy numbers. Kandala and Prajneshu (2004) have applied this methodology for fitting well-known “Allometric model” to length-weight data of some fish species. However, a drawback of this procedure is that the spread of estimated responses increases as magnitude of explanatory variable increases, even though the spread of observed responses are roughly constant or
decreasing. To overcome this, Kao and Chyu (2002) proposed a “two-stage” approach for fitting FLR model through FLS approach and showed its superiority over Diamond’s procedure. Recently, Ghosh et al. (2008) and Singh et al. (2008a) have thoroughly discussed this approach and, for its application, relevant computer programs have been developed in “Nonlinear programming solver LINGO, Version 8” software package. Specifically Possibility and Necessity measures for obtaining reliable fuzzy estimate of crop yield by estimating parameters using “Fuzzy Least squares” is carried out. As an illustration, the methodology is applied to Pearl Millet crop yield data in order to build block level estimates for Bhiwani district, Haryana based on farmers’ estimates. Performance evaluation criterion is used to compare results of Possibility, Necessity and Minimization approaches at optimum value of fitness level. Finally, fitting of fuzzy von Bertalanffy growth model, when response variable is reported in intervals corresponding to various crisp values of explanatory variable, is carried out for pearl oyster age-length data. Yang and Liu (2003) developed robust algorithms against presence of outliers in a FLR model. However, this methodology still needs to be applied to data from the field of agriculture.

4. Fuzzy nonlinear regression methodology: In contrast to fuzzy linear models, very little research work so far has been done dealing with “Fuzzy nonlinear models (FNMs)”. Buckley and Feuring (2000) proposed “Evolutionary algorithm solutions” for fitting some particular parametric FNMs. Specifically, for given fuzzy data, the algorithm searches from the “Library” of fuzzy functions (which includes linear, polynomial, exponential, and logarithmic) that function which best fits the data. Evidently, this is not at all satisfactory for fitting parametric FNMs to data and so is a fertile possible area for future research.

Hong and Hwang (2003) introduced the use of “Support vector machines” (SVM) for fitting nonparametric FNMs. SVM was developed at AT&T Bell Laboratories by V. Vapnik and his coworkers (Vapnik, 2000) and is based on the idea of Structural risk minimization. An excellent description of various aspects of SVM is given in Suykens et al. (2002). However, a lot of effort still needs to be expanded to fine-tune the methodology in order to apply it to real data.

REFERENCES


