A Frame Work for the Development of a Suitable Method to Find Shoot Length at Maturity of Mustard Plant Using Soft Computing Model

Satyendra Nath Mandal, J. Pal Choudhury, Dilip De, and S. R. Bhadra Chaudhuri

Abstract—The production of a plant can be measured in terms of seeds. The generation of seeds plays a critical role in our social and daily life. The fruit production which generates seeds, depends on the various parameters of the plant, such as shoot length, leaf number, root length, root number, etc. When the plant is growing, some leaves may be lost and some new leaves may appear. It is very difficult to use the number of leaves of the tree to calculate the growth of the plant. It is also cumbersome to measure the number of roots and length of growth of root in several time instances continuously after certain initial period of time, because roots grow deeper and deeper under ground in course of time. On the contrary, the shoot length of the tree grows in course of time which can be measured in different time instances. So the growth of the plant can be measured using the data of shoot length which are measured at different time instances after plantation. The environmental parameters like temperature, rain fall, humidity and pollution are also play some role in production of yield. The soil, crop and distance management are taken care to produce maximum amount of yields of plant. The data of the growth of shoot length of some mustard plant at the initial stage (7, 14, 21 & 28 days after plantation) is available from the statistical survey by a group of scientists under the supervision of Prof. Dilip De. In this paper, initial shoot length of Ken (one type of mustard plant) has been used as an initial data. The statistical models, the methods of fuzzy logic and neural network have been tested on this mustard plant and based on error analysis (calculation of average error) that model with minimum error has been selected and can be used for the assessment of shoot length at maturity. Finally, all these methods have been tested with other type of mustard plants and the particular soft computing model with the minimum error of all types has been selected for calculating the predicted data of growth of shoot length. The shoot length at the stage of maturity of all types of mustard plants has been calculated using the statistical method on the predicted data of shoot length.

Keywords—Fuzzy time series, neural network, forecasting error, average error.

I. INTRODUCTION

The production of a plant can be measured in terms of seeds. The generation of seeds plays a critical role in our social and daily life. The fruit production which generates seeds, depends on the various parameters of the plant, such as shoot length, leaf number, root length, root number, etc. When the plant is growing, some leaves may be lost and some new leaves may appear. It is very difficult to use the number of leaves of the tree to calculate the growth of the plant. It is also cumbersome to measure the number of roots and length of growth of root in several time instances continuously after certain initial period of time, because roots grow deeper and deeper under ground in course of time. On the contrary, the shoot length of the tree grows in course of time which can be measured in different time instances. So the growth of the plant can be measured using the data of shoot length which are measured at different time instances after plantation. The data of the growth of shoot length of some mustard plant at the initial stage (7, 14, 21 & 28 days after plantation) is available from the statistical survey by a group of scientists under the supervision of Prof. Dilip De. In this paper, initial shoot length of Ken (one type of mustard plant) has been used as an initial data.

The main work of this paper is to select a particular method which can give the best result in terms of prediction of shoot length at the stage of maturity by using certain data of initial shoot lengths after plantation of the plant. The statistical models, the methods of fuzzy logic and neural network have been tested on the mustard plant and based on error analysis (calculation of average error) that model with minimum error has been selected and can be used for the assessment of shoot length at maturity. At the beginning the data related to mustard type Ken has been used and thereafter other mustard types (B-59, Seeta, B-65, Kranti, B-54, etc.) have been used. The particular selected soft computing model with the minimum error of all types of mustard plants has been selected for calculating the predicted data of growth of shoot length. The shoot length at the stage of maturity of all types of mustard plants have been calculated using the statistical method on the predicted shoot length data computed earlier.
The objective of this work is to find out the particular plant that can grow at the desired level or not. If the plant can not be grown at desired level there is no point of maintaining the particular plant. On the contrary if the plant can grow at par with our desired level, that plant can be maintained and water and necessary fertilizer can be applied for proper growth of the plant. The methods of least square using linear, exponential, logarithmic equation and asymptotic equation have been used as statistical tools. Also, the data has been tested using fuzzy logic and neural network in the domain of soft computing.

The principal components of soft computing [24] are as follows:-

1. Neuro Computing + Fuzzy Logic (Neurofuzzy: NP)
2. Fuzzy Logic + Genetic Algorithm (Fuzzygenetic: FG)
3. Fuzzy logic + Chaos theory (fuzzychaos: FCh)
4. Neural Networks + Genetic Algorithm (Neurogenetic: NG)
5. Neural Networks + Chaos theory (Neurochaos: NCh)
6. Fuzzy Logic + Neural Networks + Genetic Algorithm (Fuzzyneurogenetic: FNG)
7. Neural Networks + Fuzzy Logic + Genetic Algorithm (Neurofuzzygenetic: NFG)
8. Fuzzy Logic + Probabilistic reasoning (Fuzzyprobabilistic: FP)


This paper presented a systematic approach to design neural networks for optimizing applications. T. K. Bhattacharya and T. K. Basu[7] described that a time series model of structural learning method was proposed which was based on interpreting hidden inputs. In order to solve these problems, a structural learning method was proposed which was based on an information criterion.

A comparative study was made among gaussian, triangular and trapezoidal functions for using the function for fuzzification[16] and it was observed that gaussian function was the most suitable function for fuzzification. A Rule Base was constructed for a Personnel Selection System using Fuzzy Expert methodology[17]. A comparison was made among the fuzzy time series and markov model [18] for the purpose of manpower prediction and it was found that fuzzy time series method was more preferable than markov model.

G. P. Bansal, A. Jain, A. K. Tiwari and P. K. Chande[19] proposed a novel methodology called genetic programming which they considered as a variant of genetic algorithms(GAs) and that evolved on a dynamic length tree representation the basis of fitness function. The methodology presented was to provide assistance to the attendant indicating that plant’s behavior, which was drifting away from normal operation with the strategy to minimize error and optimize the operation. K. K. Shukla[20] presented a novel genetically trained neural network(NN) predictor trained historical data.
Substantial improvement in prediction accuracy by the neuro-
genetic approach had been demonstrated as compared to a
regression-tree-based conventional approach and back
propagation neural network (NN) approach.

S. Bandyopadhyaya and U. Maulik[21] proposed an efficient
evolutionary search algorithm, which could exploit the
enhanced searching capability of parallel genetic algorithms. The
performance of this algorithm with and without migration
of chromosomes had been studied. In GAs (Genetic
Algorithms), all chromosomes were performing similar
operations. This made GAs (Genetic Algorithms) parallelizable relatively easily. There were advantages in parallel
GAs (Genetic Algorithms): when several GAs (genetic algorithms) were run on multiple processors simultaneously,
higher speed up might result; after some period of evolution,
the majority of chromosomes in a single population became
very similar. Depending on the parallelism, the level of
interactions, the parallel GAs (parallel Genetic Algorithms) might be
classified into two groups: (1) Fine-grain pGAs (parallel
Genetic Algorithms): Numerous small populations evolved
parallelly, with large amount of interaction among them. The
next generation was evolved by first replacing each individual
by an individual selected from its neighborhood on the grid.
Then the crossover of each individual was performed with a
mate randomly selected from its neighborhood with a certain
probability. (2) Coarse-grain pGAs (parallel Genetic Algorithms): Here small number of relatively large
populations was evolved parallelly with little amount of
interaction among them.

B. Banerjee, A. Konar and S. Mukhopadhyaya[22] proposed
a new technology for efficient obstacle avoidance and path
planning of a mobile robot by exploiting the techniques of
neural network (Hopfield network) and genetic algorithms.

After literature survey it has been observed that most of
authors[6], [8], [10], [14]) preferred the neural network
techniques for the purpose of forecasting in various
applications. In fuzzy environment, linguistic variables are
generated from crisp or real data. The states of each linguistic
variable are expressed by linguistic terms interpreted as
specific fuzzy numbers as defined in terms of a base variable,
the values of which are real numbers within a specific range.
In genetic algorithm[15], certain new strings are generated
and certain old strings are lost based on the fitness function.
Certain research work has been developed using neuro-GA
approach of software development effort[20] and for
navigational planning of a mobile robot [22] but the prediction
of shoot length at maturity was not done ever before by using
soft computing approach. That is the reason for using the
technique of soft computing approach for the purpose of
predicting the growth of shoot length in mustard type plant.

Based on the measurement of initial shoot length at the time
interval of 7, 14, 21, 28 days, the final shoot length has to be
ascertained on 95 days (day of maturity). Because the ultimate
growth of plant takes place by that maturity time and at that
time the plant can produce maximum seeds.

II. METHODOLOGY

A. Fuzzy Time Series

The main difference between fuzzy time series and
conventional time series is that the values of the former are
fuzzy sets while the values of the latter are real numbers[11 –
[3], [5]). A fuzzy set is a class with fuzzy boundaries. Let U
be the universe of discourse and U = {u1, u2, ......, un}. A
fuzzy set A of U is defined by

\[ A = \sum f_A(u_i) / u_i \]

where \( f_A(u_i) \) indicates the grade of membership of \( u_i \) in \( A \), where \( f_A(u_i) \in [0, 1] \) and \( 1 <= i <= n \).

B. Artificial Neural Network (ANN)

An ANN (Artificial Neural Network) is composed of basic
units called artificial neurons on neurodes that are processing
elements (PEs) in a network. Each neurode receives input data,
processes it and delivers a single output. The input data can be
raw data or output of other processing elements (PEs). The
output can be the final product or it can be an input to another
neurode.

An ANN (Artificial Neural Network) is composed of
collection of interconnected neurons that are often grouped in
layers; however in general no specific architecture should be
assumed. In terms of layered architecture, two basic structures
are considered. In one type, two layers are seen: input and
output. In other type there are three layers: input, intermediate(called hidden) and output. An input layer
receives data from the outside world and sends signals to
subsequent layers. The outside layer interprets signals from
the previous layer to produce a result that is transmitted to the
outside world. In three layer ANN ((Artificial Neural Network)
architecture, the concept of hidden layer is assumed in order
to control the weights between input to hidden layer and
hidden layer to output layer.

C. Feed Forward Back Propagation Neural Network

The feed forward back propagation neural network (FFBP
NN) has been used under neural network method. The feed
forward back propagation neural network (FFBP NN) does not
have feedback connections, but errors are back propagated
during training. Errors in the output determine measures of
hidden layer output errors, which are used as a basis for
adjustment of connection weights between the input and
hidden layers. Adjusting the two sets of weights between the
pairs of layers and recalculating the outputs is an iterative
process that is carried on until the errors fall below a tolerance
level. Learning rate parameters scale the adjustments to
weights. A momentum parameter can be used in scaling the
adjustments from a previous iteration and adding to the
adjustments in the current iteration. The layout of feed
forward back propagation neural network is furnished in Fig.
1.

D. Error Analysis

Forecasting error = | Forecasted Value - Actual Value | / (Actual Value) * 100%
Average Forecasting error = (Sum of Forecasting errors) / (Total no of Errors).

E. Data used in this Paper (Time Instances vs. Shoot Length of Different Mustard Plant)

A Statistical survey has been conducted by a group of certain agricultural scientists on different mustard plants under the supervision of Prof. Dilip De, Bidhan Chandra Krishi Viswavidayalay West Bengal, India. The objective of the survey is to find out the production of a particular type plant using certain initial parameters (root length, root number, leaf number, shoot length, branch number etc.). In this paper, the value of shoot length measured at initial stage after plantation has been taken as input data of a particular type of mustard plant. The value of shoot length (measured with equal time intervals within 28 days after plantation) for different mustard plants(B-59, Seeta, B-65, Kranti, B-54, etc.) have been taken and furnished in Table I.

III. IMPLEMENTATION

A. Fuzzy Time Series

In this paper, the effort has made on the initial shoot length of Ken type mustard plant and then same methods have been used all other type of mustard plants.

Step 1.

The universe of discourse U is defined within which the available data (shoot length). The minimum data (Dmin say,) is 17 and the maximum data (Dmax say,) is 67. The universe U is defined as [Dmin - D1, Dmax + D2 ] where D1 and D2 are two proper positive numbers. Let us choose D1 = 7 and D2 = 3. Thus the universe is the interval of U [10, 70].

Step 2.

The universe U is partitioned into five equal length(10) intervals. The intervals are chosen as u1 = [10, 20], u2 = [21, 30], u3 = [31, 40], u4 = [41, 50], u5 = [51, 60], u6 = [61, 70].

Step 3.

The available data are fuzzified based on gaussian function which is furnished in Table II given below:-

<table>
<thead>
<tr>
<th>Actual Shoot Length</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>1.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>22.05</td>
<td>0.8</td>
<td>1.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>27</td>
<td>0.3</td>
<td>1.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>31</td>
<td>0.0</td>
<td>0.9</td>
<td>1.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>36</td>
<td>0.0</td>
<td>0.4</td>
<td>1.0</td>
<td>0.6</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>41</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>1.0</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>45</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.5</td>
<td>0.0</td>
</tr>
<tr>
<td>49</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.0</td>
</tr>
<tr>
<td>54</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>1.0</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>57</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
<td>1.0</td>
<td>0.7</td>
</tr>
<tr>
<td>60</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>64</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>1.0</td>
</tr>
<tr>
<td>67</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.3</td>
<td>1.0</td>
</tr>
</tbody>
</table>

TABLE II

Actual Shoot Length  Fuzzy Set

It is to note that the repeated relationships are counted for only once.

Step 5.

By definition a time-invariant fuzzy time series has been developed([1], [3], [5]). Let us define an operator ’ X ’ of two vectors. Suppose C and B are row vectors of dimension m and D = (dij) = C T X B . Then the element dij of matrix D at row i and column j is defined as: d ij = min (Ci, Bj ) (i, j = 1, ........., m) where Ci and Bj are the i-th and the j-th element of C and B respectively.

Let R1 = A1T X A2, R2 = A2T X A2, R3 = A2T X A3, R4 = A3T X A3, R5 = A3T X A4, R6 = A4T X A4, R7 = A4T X A5, R8 = A5T X A5, R9 = A5T X A6, R10 = A6T X A6.

Now a relation matrix has been formed as R( t, t - 1) = R = U Ri where R is a 6 by 6 matrix and U is the union operator. The value of R is follows:-

0.8 1.0 0.7 0.1 0.0 0.0
0.7 1.0 0.6 0.1 0.0
0.2 0.7 1.0 0.5 0.1
0.0 0.1 0.6 1.0 0.6
0.0 0.0 0.1 0.6 1.0
0.0 0.0 0.0 0.3 1.0

Using R, the forecasting model is defined as,

Ai = Ai - 1 . R (2)

where Ai - 1 is the shoot length of reading i - 1 and Ai is the predicted shoot length of reading i in terms of fuzzy sets and ’ . ’ is the max - min operator.

Step 6.

The forecasted output is interpreted. The calculation results using (2) are actually all fuzzy sets. Now it is necessary to translate the fuzzy output into a regular number (equivalent
scal). This step is called defuzzification. The following principles are used to interpret the forecasting results:-

(1) If the membership of an output has only one maximum, the midpoint of the interval corresponding to the maximum is selected as the forecasted value.

(2) If the membership of an output has two or more consecutive maximums, the midpoint of the corresponding conjunct intervals is selected as the forecasted value.

(3) Otherwise, the fuzzy output is standardized and the midpoint of each interval is used to calculate the centroid of the fuzzy set as the forecasted value.

Method 1 and 2 are very popular, but do not give satisfactory results. The method 3 has been used here. The predicted values of the shoot length are furnished in Table IV.

![Table IV](image)

Step 7.

Finally the forecasted error and average forecasting error have been calculated and furnished in Table V.

![Table V](image)

B. Artificial Neural Network

The steps narrated in Step-1, Step-2 and Step-3 under fuzzy time series have been used and the fuzzy data is fed to artificial neural network.

Under artificial neural network system, a feed forward back propagation neural network is used which comprises of a 6noded input layer, 6 noded output layer and 2 noded hidden layer.

The value of the artificial neural network parameters are furnished in Table VI. It is to mention the $M_1$ array be the matrix of weights from input to the hidden layer, $M_2$ array be the matrix of weights from hidden to output layer, $A$ array be the threshold value for jth hidden layer, $K$ array be the threshold or bias to jth output layer as furnished in Table VI. The value of learning rate $B_l$ is taken as 1.5 and $B_h$ as 2.0. The momentum parameter $\alpha$ is taken as 0.7.

![Table VI](image)

Step 6.

The forecasted output is interpreted. The results are all fuzzy sets. Now it is defuzzified into regular number as per procedures narrated in step 6 of fuzzy time series.

The predicted values for the shoot length are calculated and are furnished in Table VII.

![Table VII](image)

IV. RESULTS

Under statistical methods the least square technique based on linear, exponential, logarithmic and asymptotic equation have been used. The average error of Ken mustard plant has been furnished as follows:-

![Table VII](image)
It has been observed that the average error based on neural network (1.11%) is minimum among all the methods. All these methods have been applied on other types of mustard plants and the value of average errors are furnished in Table IX.

Comparative study of methods based on various mustard plants of different types

From Table IX, it has been observed that the neural network gives minimum error in all types of mustard plants. Thus the estimated data based on neural network can be used for the futuristic prediction of shoot length of mustard plant at maturity. The growth of shoot length and error based on different models have been furnished in Figure 2 and Figure 3. Since the neural network can not generate futuristic terms, least square technique based on linear equation can be used for the prediction of shoot length at the stage of maturity (95th day after plantation). The predicted shoot length at maturity for different mustard plants have been furnished in Table X.

V. CONCLUSION

Fuzzy time series method is unsupervised model and neural network is supervised model where training of neural network exists and that has improved the performance. In case of statistical models, the slight deviation of actual data value due to certain instability of system, will generate a large amount of error where as the fuzzy logic takes care of that situation That is the reason for using neural network model using fuzzy data. Through data analysis of various mustard plants the same picture has been observed when the growth of shoot length has been used as a data.
TABLE I
SHOOT LENGTH FOR DIFFERENT MUSTARD PLANTS

<table>
<thead>
<tr>
<th>Mustard Type and Shoot Length</th>
<th>Time Instances</th>
<th>Ken B-59</th>
<th>Seeta B-85</th>
<th>Kranti i B-54</th>
<th>Sananda Jhatig</th>
<th>Vardhan</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>17</td>
<td>19</td>
<td>17</td>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>22</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>27</td>
<td>28</td>
<td>26</td>
<td>27</td>
<td>26</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>31</td>
<td>33</td>
<td>29</td>
<td>32</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>36</td>
<td>37</td>
<td>32</td>
<td>37</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>41</td>
<td>41</td>
<td>35</td>
<td>39</td>
<td>32</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>45</td>
<td>45</td>
<td>38</td>
<td>41</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>49</td>
<td>49</td>
<td>41</td>
<td>42</td>
<td>36</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>54</td>
<td>54</td>
<td>44</td>
<td>44</td>
<td>37</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>57</td>
<td>57</td>
<td>49</td>
<td>52</td>
<td>42</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>60</td>
<td>59</td>
<td>55</td>
<td>59</td>
<td>47</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>64</td>
<td>63</td>
<td>61</td>
<td>68</td>
<td>51</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>67</td>
<td>66</td>
<td>66</td>
<td>76</td>
<td>56</td>
</tr>
</tbody>
</table>

TABLE IX
COMPARATIVE STUDY OF METHODS BASED ON VARIOUS MUSTARD PLANTS OF DIFFERENT TYPES

<table>
<thead>
<tr>
<th>Methods</th>
<th>Mustard Plant Type</th>
<th>Ken B-59</th>
<th>Seeta B-85</th>
<th>Kranti i B-54</th>
<th>Sananda Jhatig</th>
<th>Vardhan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td></td>
<td>2.88</td>
<td>3.66</td>
<td>4.64</td>
<td>6.619</td>
<td>5.91</td>
</tr>
<tr>
<td>Exponential</td>
<td></td>
<td>8.83</td>
<td>7.52</td>
<td>2.52</td>
<td>7.23</td>
<td>2.7</td>
</tr>
<tr>
<td>Logarithmic</td>
<td></td>
<td>4.38</td>
<td>3.73</td>
<td>3.19</td>
<td>7.49</td>
<td>4.27</td>
</tr>
<tr>
<td>Asymptotic</td>
<td></td>
<td>3.13</td>
<td>2.44</td>
<td>9.42</td>
<td>14.27</td>
<td>7.87</td>
</tr>
<tr>
<td>Fuzzy Time Series</td>
<td></td>
<td>5.83</td>
<td>4.98</td>
<td>8.007</td>
<td>8.35</td>
<td>8.58</td>
</tr>
<tr>
<td>ANN with Fuzzy input</td>
<td></td>
<td>1.11</td>
<td>1.008</td>
<td>0.86</td>
<td>1.15</td>
<td>1.02</td>
</tr>
</tbody>
</table>

REFERENCES


