

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/267574862>

# Fuzzy Inference Systems for Crop Yield Prediction

Article in *Journal of Intelligent Systems* · January 2012

DOI: 10.1515/jisys-2012-0016

---

CITATIONS

3

---

READS

813

2 authors, including:



Jayaram M.A

Siddaganga Institute of Technology

120 PUBLICATIONS 747 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



1. Software effort estimation using soft computing techniques. [View project](#)

## Fuzzy Inference Systems for Crop Yield Prediction

M. A. Jayaram and Netra Marad

**Abstract.** Prediction of crop yield is significant in order to accurately meet market requirements and proper administration of agricultural activities directed towards enhancement in yield. Several parameters such as weather, pests, biophysical and physio morphological features merit their consideration while determining the yield. However, these parameters are uncertain in their nature, thus making the determined amount of yield to be approximate. It is exactly here that the fuzzy logic comes into play. This paper elaborates an attempt to develop fuzzy inference systems for crop yield prediction. Physio morphological features of Sorghum were considered. A huge database (around 1000 records) of physio morphological features such as days of 50 percent flowering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries and the corresponding yield were considered for the development of the model. In order to find out the sensitivity of parameters, one-to-one, two-to-one and three-to-one combinations of input and output were considered. The results have clearly shown that panicle length contributes for the yield as the lone parameter with almost one-to-one matching between predicted yield and actual value while panicle length and panicle weight in combination seemed to play a decisive role in contributing for the yield with the prediction accuracy reflected by very low RMS value.

**Keywords.** Approximate Reasoning, Crop Yield, Fuzzy Logic, Fuzzy Rule Base, Membership Functions, Physio-Morphological Features.

**2010 Mathematics Subject Classification.** 68T10.

### 1 Introduction

Sorghum [*Sorghum biocolor* (L.) Moench] is one of the major cereal crops after rice, wheat and maize. Sorghum originating in tropical Africa is a crop with extreme genetic diversity. Sorghum is an important staple food for more than 300 million people and feed for cattle living in Asia and Africa. Among the cereals, sorghum ranks fifth in the world in production next only to wheat, maize, rice and barley. India has the largest share (32.3%) of world's sorghum area and ranks second in the production after USA. World's sorghum production of 58.8 million tons comes from an area of 42.07 million ha with productivity of 1394 kg per ha [2]. In India, sorghum is cultivated in all three seasons, that is, *kharif*, *rabi* and summer. The production is around 7.00 million tonnes from an area of 9.69 million ha with

an average productivity of 733 kg per ha, which is much lower than world average (1394 kg/ha). The *Rabi* sorghum accounts for 45 percent of the total area under cultivation and 32% of the total production. Although, *kharif* yields are higher (1100 kg/ha) than *rabi* (750 kg/ha), *Rabi* sorghum is highly valued because of its excellent grain and fodder quality [3].

Accurate estimates of crop yield are becoming increasingly important in developing countries and have sustained importance in developed countries [4]. Reliable and timely forecasts provide important and useful input for proper, foresighted and informed planning, more so, in agriculture which is full of uncertainties. Agriculture nowadays has become highly input and cost intensive. Under the changed scenario today, forecasting of various aspects relating to agriculture is becoming essential. But in spite of a strong need for a reliable and timely forecast, the current status is far from satisfactory [1]. Forecasting of crop yield well before harvest is crucial especially in regions characterized by climatic uncertainties. This enables planners and decision makers to predict how much to import in case of shortfall or, optionally, to export in case of surplus [11].

The synergy of neural network, fuzzy sets and genetic algorithms is collectively known as computational intelligence. Several different combinations exist. There exist the optimization of neural network weights via a genetic algorithm and the use of a genetic algorithm to tune a fuzzy inference system. In this paper we focus on the fuzzy inference system and fuzzy set theory.

The rest of the paper is organized as follows. In Section 2, the literature review is presented and Section 3 explains the architecture of the proposed fuzzy inference system, followed by Section 4 where the data and parameters used in the model have been elucidated. Discussions are made on the results obtained in Section 5, and the paper concludes in Section 6.

## 2 Literature Review

Many researchers have attempted to develop yield prediction models in the past. Martin [7] developed ANN models using GA-selected inputs for the prediction of maize yield in the southeastern US based on indices of large-scale meteorological phenomena. In his work, ANN models are developed using inputs from four indices of meteorological phenomena. In order to determine the subset of these inputs, which would produce ANN models with the lowest error, a GA is used to search the input space.

Agarwal and Mehta [1] made weather based models for forecasting crop yield for various crops at selected states and climatic zones. The models utilize fortnightly weather data and agricultural inputs. The authors used regression analysis and developed complex polynomials using artificial neural network techniques. It

is reported that these models have been successfully used by other workers and organizations for forecasting crop yield and enabled the farmers to optimize plant protection measures.

Stathakis et al. [13] used neuro-fuzzy models for predicting the crop yield. Their work explores the dynamics of neural networks using remote sensing and other data. It is reported that the inputs for the adaptive neuro-fuzzy inference system (ANFIS) included temperature,  $CO_2$ , vapour pressure deficit, and radiation. The output is a single number, which is the yield of the crop. Then the system was evaluated by finding deviation of estimated yield and the actual yield.

Qaddoum et al. [9] also developed the adaptive neuro-fuzzy models for predicting the yield of the tomato. They used environmental factors as the input for the system. In this case also the average error in prediction was found to be low.

Sayari et al. [12] came up with the statistical model in predicting the crop yield based on the historical data. The inputs for the model were climatic parameters and the aim was to find the relationship between the crop yield and evapotranspiration.

Lobell and Burke [6] developed statistical models for predicting the potential effects of climate change on crop yield. The crop considered was maize with real field data from nearly two hundred sites in sub-Saharan Africa. They used statistical models for the predictions.

Prasad et al. [8] predicted crop yield using the soil surface parameters as the input. A non-linear quasi-Newton multi-variant optimization method was used to minimize inconsistency and errors in yield prediction. The crops considered for their model were corn and soybean in Iowa State.

The main motivation for the use of a fuzzy inference system for crop yield estimate is that fuzzy inference systems work on approximate reasoning and the capability of fuzzy based models in capturing imprecision and uncertainties of non-statistical kind. Further, due to ambiguities in the roles of several dependent parameters in controlling the yield of a crop, and also due to uncertainties and degrees of accuracies in their measurement, it is impossible to develop a crisp model that would decisively predict precise values of yield for a given set of input parameters. The ambiguity also stems from measurement error and generalization. By using fuzzy sets instead of the actual values as inputs, we have aimed at shifting to the semantics of the data rather than its measure.

Though there are quite a good number of prediction exercises reported in the literature, almost all of them are based on theory of statistics. No attempts seem to have been made in applying fuzzy reasoning in building input–output mappings related to crop yield estimation. Therefore the intent of this work was set to apply the fuzzy inference system architecture for crop yield estimation. In this context, the work is a significant contribution because the fuzziness involved in the input parameters has been captured in a more natural way.

### 3 Fuzzy Inference System Architecture

Fuzzy inference systems (FIS) are one of the most famous applications of fuzzy logic and fuzzy set theory. The strength of FIS relies on their two-fold identity. On the one hand, they are able to handle linguistic concepts. On the other hand, they are the universal approximators able to perform non-linear mappings between inputs and outputs [5].

Fuzzy rule-based systems use linguistic variables to reason using a series of logical rules that contain IF-THEN rules, which connect antecedent(s) and consequent(s), respectively. An antecedent is a fuzzy clause with certain degree of membership (between 0 and 1). Fuzzy rules can have multiple antecedents connected with AND or OR operators, where all parts are calculated simultaneously and resolved into a single number. Consequents can also be comprised of multiple parts, which are then aggregated into a single output of a fuzzy set.

The Mamdani inferencing method was used. Based on the Mamdani implication method for the set of rules the aggregated output is given by

$$\mu_B(y) = \max[\min[\mu_{A1}^k(\text{input}(i)), \mu_{A2}^k(\text{input}(j))]], \quad k = 1, 2, \dots, r \quad (1)$$

(see [10]), where  $\mu_B$  is a consequent membership function and  $A1, A2, \dots$  refer to the antecedents of the  $k$ th rule. This equation implies that the maximum membership value among the minimum values need to be propagated in order to obtain the portion of the consequent.

The inputs considered for the model are days of 50 percent flowering (%), dead heart percentage (%), plant height (cm), panicle length (cm), panicle weight (gm), and number of primaries (per panicle). Therefore, it is a six-to-one fuzzy mapping.

### 4 The Model

The data for the model relates to extensive filed research work pursued at the University of Agricultural Sciences, Dharwad, India [3]. The data set consisted of more than 1000 records each for Sorghum with six physio-morphological attributes and the corresponding yield. First, some records were removed for incompatibility and missing data, then the parameters were considered for fuzzification.

The model was developed on the modular basis considering one-to-one mappings for all the input parameters and the combination of parameters were considered for developing two-to-one, three-to-one and so on. In case of one-to-one mapping, parameters were considered on individual basis in the first attempt. As an example, in case of panicle length as an input parameter ten fuzzy gradations were made between lower bound and upper bound of the input. The term set for

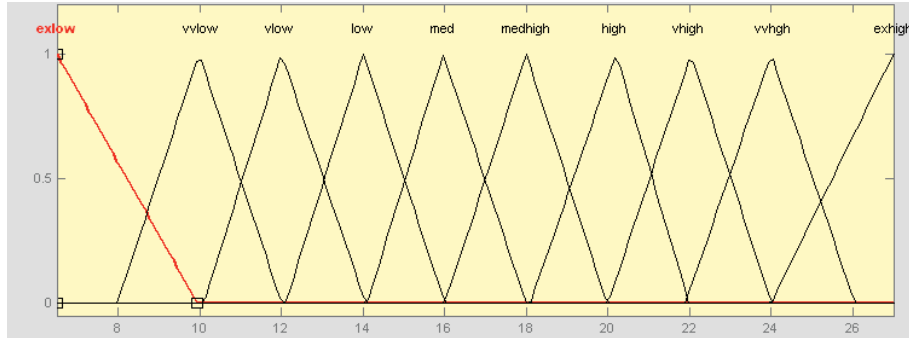


Figure 1. Fuzzy input membership functions for the parameter-panicle length (cm).

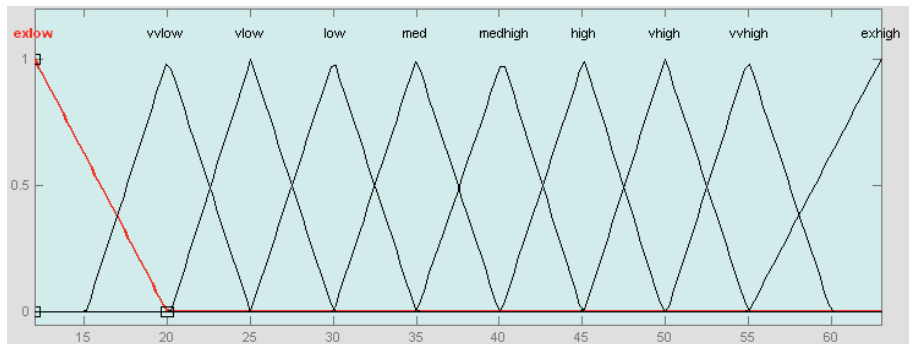


Figure 2. Fuzzy output membership functions of the crop yield.

this input is

$$PL = \{ \text{extremely low, very very low, very low, low, medium,} \\ \text{medium high, high, very high, very very high, extremely high} \}. \quad (2)$$

Figure 1 shows, the triangular membership functions for the above input parameter.

The corresponding output membership functions were also ten in number with same linguistic names as that of inputs. Figure 2 shows the output membership functions for the panicle length as input.

The fuzzy rule base consisted of ten rules. A typical rule base is shown in Figure 3.

Similar one-to-one mappings were done for the remaining five input parameters and an appropriate number of output membership functions and fuzzy rules were constructed in each case. As a sequel, pairs of inputs were considered based on

If panicle length is low, then the yield is low.  
 If panicle length is low, then the yield is medium.  
 If panicle length is medium, then the yield is medium.  
 If panicle length is high, then the yield is high.  
 If panicle length is very high, then the yield is very high.

Figure 3. A typical rule base for one-to-one mapping.

their mutual inter-relationships. It was possible to make three such combinations, namely *panicle length – panicle weight*, *plant height – panicle length*, and *panicle length – number of primaries per panicle*. The output membership functions were initially constructed based on the input output data pattern. On subsequent training, the output membership functions were fine-tuned. Similarly, three-to-one mappings were tried with appropriate combinations, it was possible to find only one such pairings of triplets. Further, four-to-one mappings were also attempted, but in vain.

## 5 Results and Discussions

The modeling involved intensive computational experiments. Among 1000 records availed, 70% of the data were used to train the system and the remaining 30% were used for testing the system prediction. The results in almost all the cases were found to be highly correlating and matching with the actual values. The prediction accuracy as reflected by RMS value ranged between 1.39 and 3.9. In comparison with the magnitude of real data values which were in the range of 14 to 50 metric tonnes/plot, the RMS value seemed to be very less as it conveys slight deviation in predictions. Table 1 shows the results of one-to-one mapping. All the six parameters, namely days of 50% flowering, dead heart percentage, plant height, panicle length, panicle weight and number of primaries, showed their significance in terms of yield of crop. The predicted values matched the actual yield values with low variation. However, the panicle length could be singled out as a highly sensitive parameter as it showed very high accuracy in predicted values with the lowest RMS value of 1.39.

Table 2 presents the results of the two-to-one mapping, FIS were framed considering several combinations, only three combinations, i.e., *panicle length – panicle weight*, *panicle length – number of primaries* and *plant height – panicle length*, showed to be appropriate duals. Here again, the prediction accuracy was quite good with RMS value between 3.09 and 3.61. Among these three combinations, *panicle length – panicle weight* outsmarted other combinations with a very low prediction error of 3.1. Other combinations tried were found to be unworthy be-

Input	Root mean square error (RMS)
Days of 50 percent flowering	2.98
Plant height	2.60
Panicle length	1.39
Panicle weight	2.96
Number of primaries	2.53

Table 1. Prediction results of one-to-one mappings.

Parameter 1	Parameter 2	Root mean square error (RMS)
Panicle length	Panicle weight	3.09
Panicle length	Number of primaries	3.34
Plant height	Panicle length	3.61

Table 2. Prediction results of two-to-one mappings.

Inputs	Root mean square error (RMS)
Panicle weight, panicle length, number of primaries	3.9

Table 3. Prediction results of three-to-one mappings.

cause of very high magnitude of errors in predictions. Several three-to-one mappings were also tried but none of the input combinations except *panicle weight*, *panicle length* and *number of primaries* was found to augur well. This combination showed a low prediction error of 3.9 as presented in Table 3.

Figure 4 shows the graphical representation of compatibility between predicted value and actual value of yield as far as panicle length (as lone parameter) is considered. Logarithm regression was found to be the best fit. Figure 5 shows the prediction behavior of the system when two inputs are considered, here again, logarithmic curve was in good agreement. Figure 6 shows a 3D view of the variation of yield when three input parameters (*panicle length*, *panicle weight*, *number of*



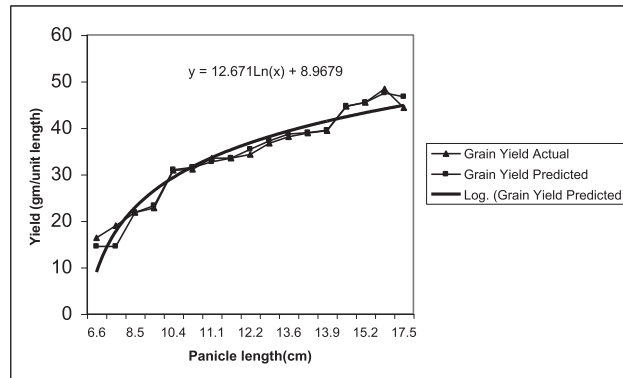


Figure 4. Predicted yield vs. actual yield with panicle length as input.

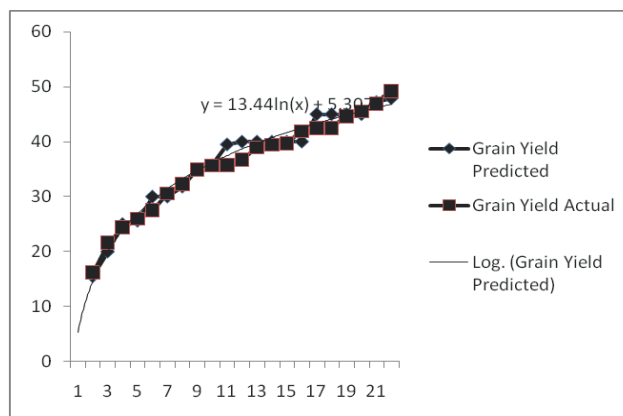


Figure 5. Predicted yield vs. actual yield with panicle length and panicle weight as inputs.

*primaries per panicle*) are considered. The high magnitude of errors in the predicted values were noticed for other combinations (four-to-one, five-to-one, six-to-one) of input–output. Such high errors may be attributed to non-interrelationships among the input parameters when they are considered in combinations.

## 6 Conclusions

This paper presented the development of a fuzzy inference system for the prediction of the yield of Sorghum crop. In a nutshell, the results emphasized three

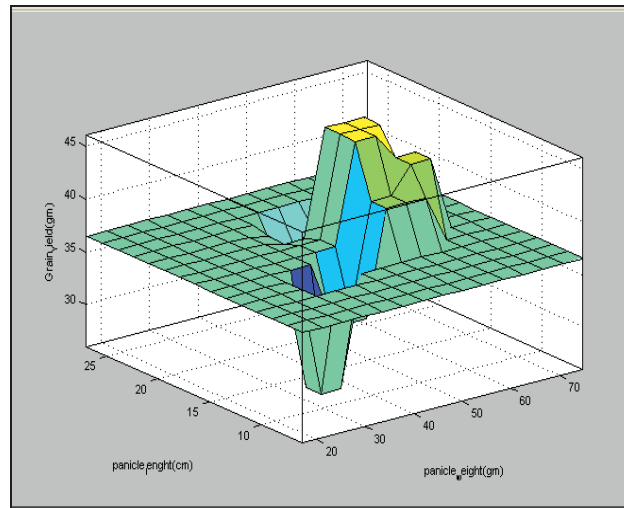


Figure 6. 3D view of three-to-one mapping (panicle length, panicle weight, number of primaries).

important points that the users of crop prediction models should consider in future work. Firstly, fuzzy based models represent a very useful tool for projecting physio-morphological features as the inputs. Second, the relative performance of the model depends on the quantum of the realistic data in hand because of approximations and uncertainties involved. Finally, the accuracy of the model depends on the granularity with which the input fuzzy membership functions are constructed.

**Acknowledgments.** The authors wish to sincerely acknowledge the help received from the University of Agricultural Sciences, Dharwad, India, while procuring research data.

## Bibliography

- [1] R. Agarwal and S. C. Mehta, Weather based forecasting of crop yields, pests and diseases – IASRI models, *Journal of Indian Society of Agricultural Statistics* **61** (2007), 255–263.
- [2] Anonymous, The state of food and agriculture, *FAO Bulletin of Statistics* **35** (2004), 38–44.
- [3] R. B. Chittapur, *Genetic studies on grain quality and productivity traits in Rabi Sorghum*, M.Sc. thesis, University of Agricultural Sciences, Dharwad, 2006.

- [4] N. G. Garcia, *Estimating maize grain yield from crop biophysical parameters using remote sensing*, Ph.D. thesis, University of Nebraska-Lincoln, 2010.
- [5] M. A. Jayaram, M. C. Nataraja and C. N. Ravikumar, Prediction of early strength of concrete: A fuzzy inference system model, *International Journal of Physical Sciences* **1** (2006), 47–56.
- [6] D. B. Lobell and M. B. Burke, On the use of statistical models to predict crop yield responses to climate change, *Agricultural and Forest Meteorology* **150** (2010), 1–10.
- [7] C. M. Martin, *Crop yield prediction using artificial neural networks and genetic algorithms*, M.Sc. thesis, University of Athens, Georgia, 2007.
- [8] A. K. Prasad, L. Chai, R. P. Singh and M. Kafatos, Crop yield estimation model for Iowa using remote sensing and surface parameters, *International Journal of Applied Earth Observation and Geoinformation* **8** (2006), 26–33.
- [9] K. Qaddoum, E. Hines and D. Illiescu, Adaptive neuro-fuzzy modeling for crop yield prediction, in: *Proceedings of the 10th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases* (Cambridge 2011), 199–204.
- [10] T. J. Ross, *Fuzzy Logic with Engineering Applications*, John Wiley and Sons, 2011.
- [11] H. L. A. Sawasawa, *Crop yield estimation: Integrating RS, GIS and management factors*, M.Sc. thesis, International Institute for Geo-information Science and Earth, 2003.
- [12] N. Sayari, M. Bannayan, A. Farid, A. Alizadeh and M. R. H. Kermani, Crop water consumption and crop yield prediction under climate change conditions at north-east of Iran, in: *International Conference on Environmental and Computer Science, IPCBEE 19*, IACSIT Press, Singapore (2011).
- [13] D. Stathakis, I. Savin and T. Nègre, Neuro-fuzzy modeling for crop yield prediction, *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **34** (2010), 105–108.

Received July 14, 2012.

#### Author information

M. A. Jayaram, Department of Master of Computer Applications,  
Siddaganga Institute of Technology, Tumkur, India.  
E-mail: jayaram\_mca@sit.ac.in

Netra Marad, Department of Master of Computer Applications,  
Siddaganga Institute of Technology, Tumkur, India.  
E-mail: netra\_marad@gmail.com