



Improvement of the Model for Predicting Rainfall by Fuzzy Set Theory Using USDA Scan Data



Hasan, M.*, Tsegaye, D.T.*, Shi, F.*, Schaefer, L.G.§, Taylor, H.G. §§.

*Alabama A & M University, Normal, AL 35762, §NRCS-National Water & Climate Center Portland, OR 97232, §§ Oregon Climate Services, Corvallis, OR 97331.

Abstract

This research presents the improvement of the fuzzy inference model primarily developed for predicting rainfall with data from United States Department of Agriculture (USDA) Soil Climate Analysis Network Station at the Alabama Agricultural and Mechanical University (AAMU) Campus for the year 2004. The primary model was developed with Fuzzy variables selected based on the degree of association of different factors with various combinations causing rainfall. An increase in wind speed (WS) and a decrease in temperature (TP) when compared between the ith and (i-1)th day were found to have a positive relation with rainfall. Results of the model showed better performance after introducing the threshold values of (i) Relative Humidity (RH) of the ith day, (ii) Humidity Increase (HI) when compared between the ith and (i-1)th day, and (iii) product (P) of increase in wind speed (WS) and decrease in temperature (TP) when compared between the ith and (i-1)th day. In case of the improved model, errors between actual and calculated amount of rainfall (RF) were 1.20, 2.19, and 9.60 percentages when using USDA scan data from AAMU campus for years 2003, 2004 and 2005, respectively. The improved model was tested at William A. Thomas Agricultural Research Station (WATARS) and Bragg farm in Alabama by checking the applicability of the model. The errors between the actual and calculated amount of rainfall (RF) were 3.20, 5.90, and 1.66 percentages using USDA scan data from WATARS for years 2003, 2004, and 2005, respectively. Similarly, percentages of errors were 10.37, 11.67, and 25.52 when using scan data from Bragg farm for years 2004, 2005, and 2006, respectively. The primary model yielded the value of error equals 12.35 percentages using USDA scan data from AAMU campus for 2004. The present model performance was proven to be better than the primary model.

Introduction

Fuzzy set theory has many potential applications to system science. Classification is a fundamental aspect of all scientific activities, and classification means putting things into sets, most of which are more meaningful in fuzzy terms than traditional crisp sets. Fuzzy set theory has acquired a misleading reputation as a new and forbidding branch of mathematics, but there is really very little that is new about it, and, far from being forbidding, many fuzzy applications are conceptually much simpler than the traditional alternatives (Silver, 2000). Weather prediction has been done using probability theory. In weather prediction, there are several factors possibly involved. An alternate method has recently been developed to describe any system that is concerned with ambiguities and vagueness, named as fuzzy set theory. Fuzzy set theory is a powerful tool and its applications have rapidly increased with establishing its utility in numerous areas of the scientific world. Any system consisting of vague and ambiguous input variables may contribute to an ultimate effect. The fuzzy logic possibility and its degree of effect due to the ambiguous input variables are considered by some as being generated in the human mind and is often referred to as expert knowledge. This expert knowledge is the accumulation of ideas as a result of the expert's experience in a particular system or system; hence, decision-making processes may be considered as fuzzy expressions perceived by the expert. Expert knowledge is expressed as a vague or ambiguous expression and not in form of any quantified value. Based on the generated idea that the possibility and degree of effect from vague and ambiguous inputs exists, then the knowledge based rules can be expressed in the form of statements, called fuzzy statements or production rules. These statements consist of antecedent (conditional part) and consequent (inference or effect due to the vague and ambiguous conditional variables) part. In predicting weather condition, there are factors in the antecedent and consequent parts are vague and ambiguous in nature. Their effect is generated in the mind of the expert with more or less accuracy. The decision-makers use their knowledge with logic and then generate algorithms in their mind (Hasan et al., 1995).

Application of the concept of fuzzy set theory in soil, crops, water management remain in its infant stages due to the lack of awareness of the superiority and potential of fuzzy set theory in the field of the above mentioned areas. Weather forecasting is one of the most imperative and demanding operational responsibilities carried out by meteorological services all over the world. It is a complicated procedure that includes numerous specialized fields of knowledge and skill. The task is intricate because in the field of agro-meteorology decisions are heavily intertwined in the visage of uncertainty associated with the weather systems.

The primary fuzzy inference model was developed successfully for predicting rainfall using the antecedent variables of increase in wind speed and decrease in temperature, hereafter termed WS and TP, respectively, when compared between the ith and (i-1)th day. Improvement of the primary model was necessary to enhance its performance in the sense of preciseness and to improve the match between actual and calculated values of rainfall, hereafter termed RF.

Modules in Fuzzy Inference Model

Fuzzification

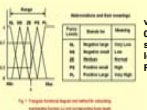
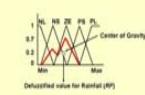


Fig. 1 shows the method of calculation for fuzzification. Here x is a value for a fuzzy variable produce two membership functions (μ) 0.3 and 0.7 against fuzzy levels PS and ZE, respectively. Table on right side shows the meanings of fuzzy levels. This method of calculating fuzzy levels and their corresponding membership functions (μ) is termed as Fuzzification.

Min-Max Composition

Fig. 2 represents the calculation method for Min-Max composition. Here, a value of WS is fuzzified which yields two membership functions (μ) 0.2 and 0.8 against fuzzy levels ZE and PS, respectively. Similarly, fuzzy variable temperature TP yields two membership functions (μ) 0.3 and 0.7 against fuzzy variables ZE and NS, respectively. Considering these two variables as the condition for rainfall RF with four fuzzy levels and their corresponding membership functions (μ) that are shown inside the blocks. According to Fig. 2, minimum value of membership function (μ) out of ZE = 0.2 (from WS) and ZE = 0.7 (from TP), 0.2 is selected as rainfall against rainfall (RF). Similarly, the minimum values of membership functions (μ) 0.7, 0.2, 0.2, and 0.3 are selected for the fuzzy levels ZE (for all the 3 blocks as they are inferred fuzzy levels for rain (RF)). Now, selecting the membership function (μ) = 0.2 for NS and the maximum value of membership function (μ) 0.7 out of 0.2, 0.3 and 0.7 for fuzzy levels ZE. Finally, two fuzzy levels NS and ZE yield two membership functions (μ) values 0.2 and 0.7, respectively. As stated here, the calculation method to yield two fuzzy levels and their corresponding membership functions (μ) values is termed as Min-Max composition.

Defuzzification



Membership functions (μ) and their corresponding fuzzy levels obtained from Min-Max composition are taken in a triangular membership functional diagram. Membership functions (μ) values 0.2 and 0.7 obtained from Min-Max composition produce a new polygon that is shown with different color under NS and ZE. Now calculate the coordinate of the center of gravity for this new polygon where x-axis gives the value for rainfall (RF). This method of calculation to yield the value for rainfall (RF) is termed as Defuzzification.

Improved Model Structure

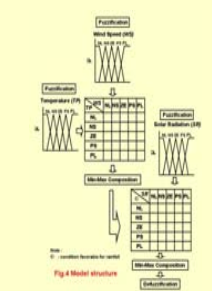


Fig. 4 illustrates the model diagram where fuzzy variables WS and TP were fuzzified. Primary Min-Max composition was done to yield values of membership functions (μ) and their corresponding fuzzy levels, and also to represent the condition (C) favorable for rainfall occurrence. Membership functions (μ) and fuzzy levels obtained from primary Min-Max composition along with the membership functions (μ) and their corresponding fuzzy levels after fuzzification for solar radiation (SR) were again considered for secondary Min-Max composition only if solar radiation (SR) value showed a decreasing trend for two consecutive days. On the other hand, if solar radiation (SR) does not have a decreasing trend for two consecutive days, only primary Min-Max composition was considered. Finally, defuzzification was done to obtain the predicted value of rainfall (RF).

Analyzing the initial results and raw data, assumptions made on relative humidity on ith day, humidity increase when compared between ith and (i-1)th day and product of (WS) and (TP) were as follows:

Threshold Boundaries

Two threshold limits, left boundary (A) and right boundary (B) were selected. Values for (A) and (B) vary depending on 3 seasons (Refer Fig. 5).

Assumptions on Relative Humidity

- If the value of relative humidity on ith day is > (B) then there is a rainfall occurrence,
- If the value of relative humidity on ith day is in between (A) and (B) then there is a possibility of rainfall occurrence,
- If the value of relative humidity on ith day is < (A) and value of solar radiation is < 40 then there is a rainfall occurrence else there is no rainfall occurrence.

Assumptions on Humidity Increase between (A) and (B)

- If the humidity increase between ith and (i-1)th day > 10 and it is beyond (B) then there is a rainfall occurrence with a fuzzy level of PS only if
 - Humidity increase between ith and (i-1)th day + Humidity > 100 and
 - (Humidity + 5) > 90.
- If Humidity increase between ith and (i-1)th day is within the range of ≤ 10 and > 10 then there is a possibility of rainfall occurrence,
- If Humidity increase between ith and (i-1)th day is ≤ 10 and solar radiation (SR) is < 40 then there is a possibility of rainfall else no rainfall.

Assumptions on product of WS and TP, between (A) and (B)

- If the product of WS and TP is ≤ 5 then there is no rainfall occurrence,
- If the product of WS and TP is ≤ 5 then there is no rainfall occurrence only if
 - WS is not increasing, TP is not decreasing and (b) solar radiation (SR) > 100 then there is a light rainfall with fuzzy level NL.

Error Calculation

Percentage of error between the actual and predicted amount of rainfall was calculated by the following equation:

$$\text{Error} = \frac{\sum_{i=1}^n \text{abs}(\text{RF}_a - \text{RF}_c)}{\sum_{i=1}^n \text{RF}_c} \times 100\%$$

Here, RF_a is the actual amount of rainfall in mm, RF_c is the predicted amount of rainfall in mm and n is the total number of rainfall occurrences.

Results and Discussions

1. Selection of variables in the model

Fuzzy variables of increase in wind speed (WS) and decrease of temperature (TP) between ith and (i-1)th day were a good choice in developing the primary model for predicting rainfall (RF). The fuzzy variables were chosen by judging the degree of involvement or association of the variables of increase in wind speed (WS) and decrease of temperature (TP) between ith and (i-1)th day against occurrence of rainfall (RF). The model also describes a secondary method how the variables can be selected without even being perceived by an expert.

2. Selection of fuzzy levels for the inference part of production rule table in the model

Selection of the fuzzy levels in the inference part of the production rules is a cumbersome process by trial and error method. The most logical fuzzy levels were chosen for rainfall (RF) and were used in the computer program. Finally the fuzzy levels for rainfall (RF) were selected based on the lowest percentage of error.

3. Maximum amount of rainfall (RF)

Real data for rainfall showed that AAMU campus weather station had only 6 actual rainfall events exceeding 50 mm in 2004. To make an uniformity in data distribution from fuzzy levels NL to PL, the following sub-ranges were used:

- first sub-range between NL and NS was taken equal to total range divided by 8 (50 ÷ 8 = 6.25),
- second sub-range between NS to ZE was taken equal to total range divided by 4 (50 ÷ 4 = 12.50),
- third sub-range between ZE to PS was taken equal to total range divided by 2 (50 ÷ 2 = 25.00), and
- fourth sub-range between ZE to PS was taken equal to total range divided by 1 (50 ÷ 1 = 50.00).

Therefore, the sub-ranges were (0 to 6.25), (6.25 to 12.50), (12.50 to 25.00), and (25.00 to 50.00) mm. This means, the bases of triangles in Fig. 2 intersect with the x-axis at 0, 6.25, 12.50, 37.50, and 50.00. In case of the primary model values of the sub-ranges were (0 to 12.50), (12.50 to 25.00), (25.00 to 37.50), and (37.50 to 50.00), respectively. Based on the results, considering the value of maximum rainfall as 50 mm was justifiable.

4. Selection of other variables and threshold values

To match the actual and predicted amount of rainfall (RF), the factors involved in the model were needed to be considered with their threshold values. These factors are

- average daily Relative Humidity,
- humidity increase between ith and (i-1)th day, and
- product of TP and WS

Introducing the threshold values decreased the percentage of error and showed a better match between actual and predicted amount of rainfall (RF).

Conclusions

Selection of variables and the fundamental logic that the values of WS and TP between the ith and (i-1)th day was a successful attempt to determine the amount of RF and its time of occurrence as the consequent part of the fuzzy inference model. Introducing the idea of threshold values of (a) RH of the ith day, (b) HI when compared between the ith and (i-1)th day, and (c) P, on the primary model and other additional considerations in the improved model matched well between the actual and predicted amount of RF and also decreased the error. The additional considerations were as follows:

- Introducing SR value of < 40 and its effect below the left boundary value of RH,
- Introducing of SR value of < 40 and its effect below the left boundary value of HI (5 ÷ 10),
- Introducing of idea "if (HI + RH of the ith day) > 100 and (RH + 5) > 90" then there will be RF of fuzzy level PS,
- Introducing the idea that "if (wind speed is not increasing or temperature is not decreasing) and SR > 100" then there will be RF with fuzzy level NL,
- Introducing the idea of primary and secondary min-max compositional approach for the condition that SR has a decreasing trend for 2 consecutive days proved to be successful.

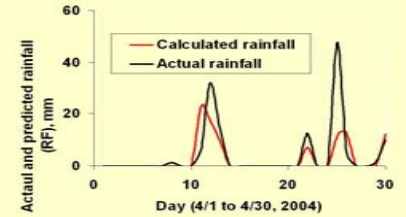


Fig. 6 Actual and predicted rainfall using USDA scan data from AAMU campus

Table 1. Results of the model when USDA data from AAMU campus

| Year | No. of actual rainy day | No. of predicted rainy day | Actual amount of Rainfall (mm) | Predicted amount of Rainfall (mm) | Error (%) |
|------|-------------------------|----------------------------|--------------------------------|-----------------------------------|-----------|
| 2003 | 91 | 91 | 1344 | 1357 | 1.20 |
| 2004 | 90 | 86 | 1465 | 1433 | 2.19 |
| 2005 | 80 | 71 | 1003 | 912 | 9.60 |

Table 2. Results of the model when USDA data from Bragg farm

| Year | No. of actual rainy day | No. of predicted rainy day | Actual amount of Rainfall (mm) | Predicted amount of Rainfall (mm) | Error (%) |
|------|-------------------------|----------------------------|--------------------------------|-----------------------------------|-----------|
| 2004 | 92 | 82 | 1505 | 1349 | 10.30 |
| 2005 | 72 | 61 | 1020 | 901 | 11.69 |
| 2006 | 76 | 43 | 1005 | 749 | 25.52 |

Table 3. Results of the model when USDA data from WATAR farm

| Year | No. of actual rainy day | No. of predicted rainy day | Actual amount of Rainfall (mm) | Predicted amount of Rainfall (mm) | Error (%) |
|------|-------------------------|----------------------------|--------------------------------|-----------------------------------|-----------|
| 2003 | 93 | 84 | 1442 | 1396 | 3.20 |
| 2004 | 80 | 87 | 1369 | 1450 | 5.90 |
| 2005 | 78 | 73 | 1021 | 1004 | 1.66 |

Acknowledgements

The authors appreciate the Faculty and Staff in the Department of Natural Resources and Environmental Sciences (NRES) of Alabama Agricultural and Mechanical University (AAMU) for their constructive suggestions and support. Thanks to Mr. Daryl Lawson, Prof. Rao Mendretrey AAMU and Dr. Rafiq Islam of Ohio State University for their constructive suggestions. United States Department of Agriculture (USDA) and Natural Resources Conservation Service (NRCS) are duty acknowledged for providing funds for conducting this research. The authors also acknowledge the late Robert Metz for providing data related to this research. The authors further extend their appreciation to (1) Dr. John A. Replogle, Research Hydraulic Engineering Collaborator, U.S. Army-Land Agricultural Research Center, USDA-ARS, (2) Dr. Rogelio C. Lazo, Free Lance International Consultant, Integrated Water Resource Planning and Management, The Philippines, and (3) Dr. O.S. Mbuya, Associate Professor, College of Engineering Sciences, Technology and Agriculture, Center for Water and Air Quality, Florida A&M University, USA.