available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/agwat

ELSEVIER

Agricultural Water Management

Model for predicting rainfall by fuzzy set theory using USDA scan data

M. Hasan[a,*], T. Tsegaye[a,1], X. Shi[b,2], G. Schaefer[c,3], G. Taylor[d,4]

[a] Dept. of Natural Resources and Environmental Sciences, Alabama Agricultural and Mechanical University (AAMU), 4900 Meridian St., Normal, AL 35762, USA

[b] Dept. of Computer Science, Alabama Agricultural and Mechanical University (AAMU), 4900 Meridian St., Normal, AL 35762, USA

[c] NRCS-National Water and Climate Center, 1201 NE Lloyd Blvd., Suite 802, Portland, OR 97232, USA

[d] Oregon Climate Service, 328 Strand Agricultural Hall, Oregon State University, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Received 30 July 2007
Accepted 27 July 2008
Published on line 23 October 2008

Keywords:
Production rules
Regression coefficient
Fuzzy inference model
Fuzzy levels
Membership function

ABSTRACT

This paper presents a fuzzy inference model for predicting rainfall using scan data from the USDA Soil Climate Analysis Network Station at Alabama Agricultural and Mechanical University (AAMU) campus for the year 2004. The model further reflects how an expert would perceive weather conditions and apply this knowledge before inferring a rainfall. Fuzzy variables were selected based on judging patterns in individual monthly graphs for 2003 and 2004 and the influence of different variables that cause rainfall. A decrease in temperature (TP) and an increase in wind speed (WS) when compared between the ith and (i − 1)th day were found to have a positive relation with a rainfall (RF) occurrence in most cases. Therefore, TP and WS were used in the antecedent part of the production rules to predict rainfall (RF). Results of the model showed better performance when threshold values for: (1) relative humidity (RH) of ith day, (2) humidity increase (HI) between the ith and (i − 1)th day, and (3) product (P) of decrease in temperature (TP) and an increase in wind speed (WS) were introduced. The percentage of error was 12.35 when compared the calculated amount of rainfall with actual amount of rainfall.

Published by Elsevier B.V.

1. Introduction

In predicting weather conditions, factors in the antecedent and consequent parts that exhibit vagueness and ambiguity are being treated with logic and valid algorithms (Hasan et al., 1995). Use of fuzzy set theory has been proved by scientists to be applicable with uncertain, vague and qualitative expressions of the system. Application of fuzzy set theory in soil, crop, and water management is still in its infant stage due to the lack of awareness of the potentials of fuzzy set theory. Weather forecasting is one of the most important and demanding operational responsibilities carried out by meteorological services worldwide. It is a complicated procedure that includes numerous specialized technological fields. The task

* Corresponding author. Tel.: +1 256 372 4560; fax: +1 256 372 8404.
E-mail addresses: mahbub.hasan@aamu.edu (M. Hasan), teferi.tsegaye@aamu.edu (T. Tsegaye), frank.shi@aamu.edu (X. Shi), garry.schaefer@por.usda.gov (G. Schaefer), taylor@coas.oregonstate.edu (G. Taylor).
[1] Tel.: +1 256 372 4219; fax: +1 256 372 8404.
[2] Tel.: +1 256 372 4144; fax: +1 256 372 5578.
[3] Tel.: +1 503 414 3068; fax: +1 503 414 3101.
[4] Tel.: +1 541 737 5705; fax: +1 541 737 5710.
0378-3774/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.agwat.2008.07.015

is complicated in the field of meteorology because all decisions are made within a visage of uncertainty associated with weather systems. Chaotic features associated with atmospheric phenomena have also attracted the attention of modern scientists. The drawback of statistical models is dependence, in most cases, upon several tacit assumptions regarding the system (Wilks, 1998). Carrano et al. (2004) compared non-linear regression modeling and fuzzy knowledge-based modeling, and explained that fuzzy models are most appropriate when subjective and qualitative data are utilized and the numbers of empirical observations are small. Brown-Brandl et al. (2003) used four modeling techniques to predict respiration rate as an indicator of stress in livestock. Four modeling techniques consisted of two multiple regression and two fuzzy inference systems. Fuzzy inference models offered better results than the two multiple regression models (Brown-Brandl et al., 2003). Fuzzy inference models yield a lower percentage of error when compared to the linear multiple regression model (Hasan et al., 1995). Similar research by Wong et al. (2003) compared the results of fuzzy rule-based rainfall prediction with an established method which uses radial basis function networks and orographic effect. They concluded that fuzzy rule-based methods could provide similar results from the established method. However, the method has an advantage of allowing the analyst to understand and interact with the model using fuzzy rules. Lee et al. (1998) considered two smaller areas where they assumed precipitation is proportional to elevation. Predictions of those two areas were made using a simple linear regression based on elevation information only. Comparison with the observed data revealed that the radial basis function (RBF) network produced better results than the linear regression models. Hence, considering the advantage of using the concept of fuzzy logic for predicting rainfall as stated by other researchers is justifiable. The advantage of fuzzy inference modeling can reflect expert knowledge and yield results with precision and accuracy. In fuzzy rule basics, knowledge acquisition is the main concern for building an expert system. Knowledge in the form of IF–THEN rules can be provided by experts or can be extracted from data. Each rule has an antecedent part and a consequent part. The antecedent part is the collection of conditions connected by AND, OR, NOT logic operators and the consequent part represents its action (Pant and Ashwagosh, 2004). In a fuzzy inference engine, the truth-value for the premise of each rule is computed and applied to the conclusion part of each rule. This result is one fuzzy subset being assigned to each output variable for each rule. For composite rules usually min–max inference technique is used.

Defuzzification is used to convert fuzzy output sets to a crisp value. The widely used methods for defuzzification are center of gravity and mean of maxima.

Generating production rules for fuzzy inference modeling is cumbersome if they are not derived as they are being perceived by an expert. Production rules have the form:

$$\text{IF } X \text{ is } A_1 \text{ AND } Y \text{ is } B_1 \text{ THEN } Z \text{ is } C_1 \tag{1}$$

$$\text{IF } X \text{ is } A_2 \text{ AND } Y \text{ is } B_2 \text{ THEN } Z \text{ is } C_2 \tag{2}$$

Here X and Y represent two antecedent variables (the conditional part of the production rule, like TP and WS as explained above), and Z is the variable yielding the consequent part of the production rule. A_1, A_2, B_1, B_2, C_1, and C_2 are the linguistic and vague expressions with ambiguities. Focusing this idea of production rule, an example for such production rule that can be employed in the present research is shown as:

IF WP is *very high* AND TP is *lower* THEN RF is *moderate* (3)

Eq. (3) shows the qualitative form of explanation, such as *very high*, *lower* and *moderate* which are all fuzzy in nature. These are explained linguistically without specific quantity or as a crisp value. The relationship of the variables between antecedent and consequent parts represent a production rule in Eq. (3) based on valid logic. In the complex reality of the world, it is usually not easy to construct rules due to the limitations of manipulation and verbalization of experts (Abe and Ming-Shong, 1995). This method is termed as the fuzzy adaptive system (FAS).

2. Definitions

2.1. Fuzzy set

Fuzzy sets are the collection of objects with the same properties, and in crisp sets the objects either belong to the set or do not. In practice, the characteristic value for an object belonging to the considered set is coded as 1 and if it is outside the set then the coding is 0. In crisp sets, there is no ambiguity or vagueness about each object belongs to the considered set. On the other hand, in daily life humans are always confronted with objects that may be similar to one other with quite different properties. Therefore, uncertainty always arises concerning the assessment of membership values 0 or 1. Logically, of course, some of the similar objects may partially belong to the same set, therefore, an ambiguity emerges in the decision of belonging or not. In order to alleviate such situations Zadeh (1965) generalized the crisp set membership degree as having any value continuously between 0 and 1. Fuzzy sets are a generalization of conventional set theory. The basic idea of fuzzy sets is easy to grasp. An object with membership function 1 belongs to the set with no doubt and those with 0 membership functions again absolutely do not belong to the set, but objects with intermediate membership functions partially belong to the same set. The greater the membership function, the more the object belongs to the set (Hasan and Zenkai, 1999).

The membership function of a fuzzy set is a generalization of the indicator function in classical sets. In fuzzy logic, it represents the degree of truth as an extension of valuation. Degrees of truth are often confused with probabilities, although they are conceptually distinct, because fuzzy truth represents membership in vaguely defined sets, not likelihood of some event or condition.

For the universe X and given the membership-degree function $\mu \rightarrow [0, 1]$ the fuzzy set is defined as

$$A = \{(x, \mu_A(x)) | x \in X\} \tag{4}$$