



OPEN ACCESS

EDITED BY

Firdous A. Shah,
University of Kashmir, India

REVIEWED BY

Mohd Tahir Ismail,
University of Science Malaysia (USM), Malaysia
Hamdi Akhsan,
Sriwijaya University, Indonesia

*CORRESPONDENCE

Ali Salman Habeeb
✉ asalhabeeb@gmail.com

RECEIVED 25 October 2025

REVISED 06 January 2026

ACCEPTED 08 January 2026

PUBLISHED 30 January 2026

CITATION

Habeeb AS, Hasan HA and Al-Sinjary AM
(2026) Fuzzy decision-tree regression model
and its application to measure some climate
change factors.
Front. Appl. Math. Stat. 12:1732313.
doi: 10.3389/fams.2026.1732313

COPYRIGHT

© 2026 Habeeb, Hasan and Al-Sinjary. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Fuzzy decision-tree regression model and its application to measure some climate change factors

Ali Salman Habeeb^{1*}, Hussein A. Hasan¹ and
Adnan M. Al-Sinjary²

¹Department of Statistics, University of Sumer, Rafai, Iraq, ²Department of Statistics and Informatics Techniques, Northern Technical University, Mosul, Iraq

In this paper, we considered a new study that examines the topic of climate change based on data from two important variables: temperature and wind speed. The study aims to employ a decision-making method based on fuzzy logic to overcome the issue of ambiguity and uncertainty. Our proposed idea in this paper was to construct an appropriate analytical framework for the phenomenon, with the aim of arriving at a more accurate decision to overcome the risks of this phenomenon and take appropriate precautions in the near and distant future to deal with this natural emergency that is increasing over time. We discussed how to implement the GUIDE regression tree algorithm as a main tool in analyzing fuzzy sets using the Triangular Membership Function to fuzzify the data to obtain more accurate partial fuzzy sets for description in the analysis of chi-square tables to make a decision using a suitable hypothesis for this purpose. The proposed method was applied to a sample size of 425 daily observations in Dhi-Qar Governorate, Iraq, for the period from December 2024 to February 2025. We used a special code in R programming for the purpose of analysis and obtaining results. Through analyzing the results, we found that two variables (temperature and wind speed) have a fundamental influence on the speed of climate change.

KEYWORDS

algorithm GUIDE, chi-square test, fuzzy decision tree, fuzzy regression, triangular membership function

1 Introduction

The problem of decision-making using statistical methods is one of the important approaches that has attracted the attention of many researchers, aiming to select the best decision by the regression model for a set of crisp data. Study problem: One of the most scientific challenges confronting decision-making is the issue of ambiguity and uncertainty. Regression trees are widely used in statistical literature as one of the most important theoretical methods employed in this area. On the other hand, one of the most important methods for dealing with data to eliminate ambiguity is fuzzy set theory, particularly when dealing with variables that can be described as categorical variables (reading them as categorical groups). Climate change phenomena are ones we encountered daily that impact our lives (temperature and wind speed... etc.). We used fuzzy set theory to represent the phenomenon under study, which has been employed to describe certain variables of the phenomenon both qualitatively and quantitatively. In this context, we focus on both aspects in order to develop an applied model that represents two types of variables

within the dataset, ultimately leading to optimal decision-making using fuzzy decision tree regression. Our interest is focused on two method types of the analytical approaches to achieve a more accurate interpretation based on a comprehensive understanding of the relationships between methods. The first involves testing the acceptance or rejection approach to analyze the significance of explanatory variables in making correct decisions, based on the appropriate tree structure for the studied phenomenon. The second focuses on modeling using regression analysis, which is based on constructing a crosstabs analysis to applying and analysis the relationship between categorical variables. Many researchers have contributed to this literature. Suarez and Lutsko [1] presented a study focused on constructing a fuzzy decision tree for regression and classification. Fuzziness is investigated by integrating fuzzy logic with decision trees of the CART type. A training rule for fuzzy decision tree was developed, resembling the backpropagation algorithm used in neural networks. This rule is compatible with a high-quality optimization algorithm designed to determine the parameters of fuzzy partitions. Wei-Yin Loh [2] presented a comprehensive review of regression and classification trees by examining several available algorithms and comparing their capabilities, strengths, and weaknesses through the application of two examples. He designed a classification tree for dependent variables that take a limited number of unordered values, where prediction error is measured by the cost of misclassification. In other words, regression trees were developed for dependent variables that take continuous or ordered discrete values, with prediction error typically measured by the squared difference between the predicted and actual values. Segatori et al. [3] proposed a distributed learning model for fuzzy decision trees based on the MapReduce framework, aimed at constructing both binary-split and multi-split trees from big data. The proposed model is based on a distributed fuzzy discretizer, which generates strong fuzzy partitions for each continuous attribute using fuzzy information entropy. These partitions are then used as inputs for the fuzzy decision tree learning algorithm, which selects the most appropriate features at decision nodes based on fuzzy information gain. The results demonstrated that the proposed approach outperforms existing methods by achieving high performance while reducing computational complexity, making it an effective solution for big data classification using distributed fuzzy decision trees. Saeed Mohammadiun et al. [4] developed a framework for designing and optimizing Fuzzy Decision Tree Regression (FDTR) models, aimed at selecting the most suitable response strategies for oil spill incidents in the harsh Arctic environment. The study employed three types of regression analysis—linear, non-linear, and Gaussian Process Regression (GPR). Additionally, four information evaluation metrics were used for decision tree splitting: information gain, deviance, GINI impurities, and misclassification error. To enhance the predictive performance of the FDTR models, the Non-dominated Self-adaptive Differential Evolution (NSDE) algorithm was applied. When tested on oil spill data, the results showed a 14% improvement in prediction accuracy and a 57% reduction in the number of rules, thereby enhancing the efficiency and robustness of the mode. Pavlos Nikolaidis [5] conducted a study based on real-world data related to energy demand and wind power generation. Regression Trees were used to forecast future

renewable energy production. The following climatic factors were used as inputs in distribution networks across different regions: wind speed and direction, ambient temperature, relative humidity, renewable energy capacity, and curtailed renewable energy output. One of the key findings of the study was that in future low-carbon energy systems, the curtailment of renewable energy production will play a significant role in intelligent forecasting systems. Therefore, accurately modeling the relationship between inputs and outputs is essential.

2 The concept of fuzzy logic

Fuzzy logic was developed in 1965 by Lotfi Zadeh, a scientist of Azerbaijani origin from the University of California, who introduced it as a better method for handling data. Fuzzy logic is a logical system based on a generalization of classical logic [6]. In other words, it encompasses theories and techniques that utilize fuzzy sets without crisp boundaries (i.e., boundaries that are unknown, undefined, or ambiguous) [7]. Fuzzy logic provides a simple way to describe and represent human expertise. Moreover, it offers practical solutions to real-world problems that are cost-effective and reasonable compared to those offered by other techniques [8].

3 The basic definitions

A classical (crisp) set is defined as a Set of elements or objects $x \in X$, which may be countable or uncountable, where each element either belongs to the set A or does not belong to it $A \subseteq X$. Hence A can be characterized by the indicator function $\mu_A(x)$ [9, 35].

Definition 1: If X is a set of objects generally denoted by x , then a fuzzy set is defined as a set \tilde{A} of ordered pairs [10]:

$$\tilde{A} = \{x, M_{\tilde{A}}(x) \mid x \in X\} \quad (1)$$

Where $M_{\tilde{A}}(x)$ is called the membership function, which also defines the degree of membership of x in the set A . When the membership function $M_{\tilde{A}}(x)$ takes only two values 0 or 1 the set A becomes a classical (non-fuzzy) set. The range of the membership function is a subset of the non-negative real numbers. Generally, elements with a membership degree of 0 are not considered part of the fuzzy set [11].

Definition 2: Let Ω be a some set A Fuzzy subset \tilde{A} of Ω define by its membership function written as $\tilde{A}(x)$ which produces values in $[0, 1]$ for all x in Ω so $\tilde{A}(x)$ is a function mapping from Ω into $[0, 1]$.

Not that if $\tilde{A}(x_0) = 1$ then we say x_0 belong to \tilde{A} and if $\tilde{A}(x_1) = 0$ we say x_1 dose not belong \tilde{A} and if $\tilde{A}(x_2) = 0.6$ we say the membership value of x_2 in \tilde{A} is 0.6 [12].

4 Fuzzy number

Definition 3: A fuzzy number is subset denoted by \tilde{x} of the set of real numbers denoted by R and is characterized by the function so called membership function $\mu_{\tilde{x}}(x)$ Fuzzy number satisfy the following constraints [13].

- (1) $\mu_{\tilde{x}} : R \rightarrow [0, 1]$ is Borel – measurable.
- (2) $\exists x_0 \in R : \mu_{\tilde{x}} = 1$.

Then \tilde{A} is called a fuzzy number and the function $\mu_{\tilde{A}}$ is called the membership function of the fuzzy number \tilde{A} [14].

5 Fuzzy number membership functions

The membership function plays a necessary role in Fuzzy Set Theory, as it constitutes one component of the ordered pair that defines a fuzzy set [15] Membership functions are used to determine the degree of membership of an element to a fuzzy set. In a fuzzy set A, an element x belongs to the set partially according to a specific membership function $\mu_A(x)$ (also referred to in some sources as fuzzification functions). The fundamental requirement for such a function are that the range of its elements is within the interval [0, 1], which determines the degree to which an element belongs to the set [16]. There are various types of membership functions, each applied to a specific phenomenon depending on its nature, where the data of the phenomenon are represented as fuzzy sets [17]. There are two main approaches for determining the appropriate membership function:

- (1) Based on human expertise: That means fuzzy sets are often used to represent and formalize human knowledge, and the membership functions constitute a part of that knowledge.
- (2) Use collected data to determine the membership function: In this approach, the structure (form) of the membership function is first specified, and then the parameters of the function are fine-tuned based on the observed data [18].
- (3) In this study, we proposed using sample quartiles to determine the parameters of the Membership functions.

Quartiles can be defined as three statistical measures that divide an ordered dataset into approximately four equal parts. Quartiles are denoted by the symbol q_i , where $i = 1, 2, 3$ [16].

q_1 : The first quartile is defined as the 25th percentile where lowest 25% data is below this point of the total data.

q_2 : The second quartile is defined as the 50th percentile where lowest 50% data is below this point of the total data, The second quartile so called the median point.

q_3 : The third quartile is the 75th percentile where lowest 75% data is below this point, It is known as the upper quartile. It can be calculated by arranging the data in ascending order, then calculating the quartile rank (quartile position) as in the Equation 2.

$$C_i = N \frac{i}{4} \tag{2}$$

Where i represents the symbol of the quartile to be calculated, C Represents the location of the quartile [19].

Membership functions take various forms, the most well-known of which include the following.

5.1 Triangular membership function

The membership values of elements belonging to a fuzzy set can be represented by a straight line, known as a linear function. This function is characterized by three main parameters (boundaries): a, b, and c. It can be defined according to the Equation 3.

$$\mu_A(x; a, b, c) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{b-a} & a < x \leq b \\ \frac{c-x}{c-b} & b < x \leq c \\ 0 & x \geq c \end{cases} \tag{3}$$

Where $a < b < c \in R$. In Figure 1a illustrates the graph of the triangular membership function (see Hasan and Mohammad [15]).

5.2 Trapezoidal membership function

It is also a linear function, and it is distinguished from other membership functions by having four parameters (boundaries). a, b, c, d This function can be defined according to the Equation 4,

$$\mu_A(x; a, b, c, d) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } b \leq x < c \\ \frac{d-x}{d-c} & \text{if } c \leq x \leq d \\ 0 & \text{if } x > d \end{cases} \tag{4}$$

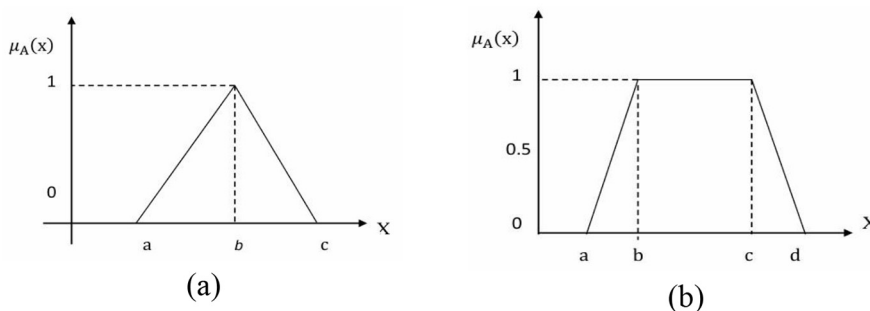


FIGURE 1 This figure represented the shapes of the membership functions: (a) first type the triangular, (b) Second type the trapezoidal.

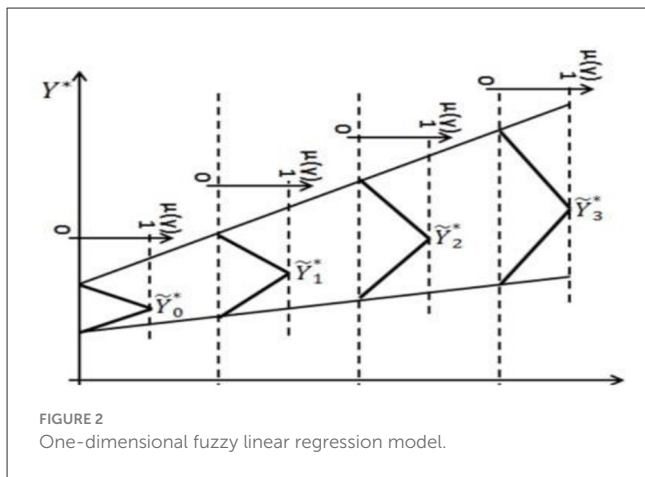


FIGURE 2 One-dimensional fuzzy linear regression model.

Where $b < c < d \in R$. In Figure 1b illustrates the graph of the trapezoidal membership function (see Alavala [20]).

6 Fuzzy linear regression model

Uncertain formulations can be used to model phenomena characterized by ambiguity through the use of fuzzy regression models, which are descriptive in nature and involve linguistic variables [21].

Figure 2 shows how the variable is divided into fuzzy categorical subdivisions, which represent the belonging functions, to show how those subdivisions overlap as linguistic variables for a single fuzzy variable in the regression model.

7 Reasons for using the fuzzy linear regression model

There are several reasons for using the linear regression model within This literature of fuzzy logic, the most important of which include the following [22]:

- (i) When the linear relationship is not well-defined.
- (ii) The assumptions of classical regression models are often strict, particularly in terms of specifying the distribution of the random error term and the relationships among explanatory variables.
- (iii) When the data exhibit an unclear or ambiguous linear trend.
- (iv) When the number of observations for the studied phenomenon is limited, providing insufficient information.
- (v) Inaccuracy in results due to uncertainty and imprecision, which are associated with vague or linguistic variables [21].

8 Description of the decision tree

One of the fundamental features of machine learning methods is the requirement of a set of numerical values known as input data.

An appropriate machine learning algorithm typically featuring a backfitting mechanism is then applied, resulting in a set of values referred to as output data [23]. The application of the decision tree regression method requires a precise description of the tree structure relevant to the study. The data are typically divided into two subsets: the training set and the testing set [24]. To understand how the decision tree operates depending on the type of tree selected various forms exist. In this study, Our interest will be on a binary-split decision tree (i.e., with left and right nodes), which is characterized by three types of nodes: a root node, internal split nodes, and terminal leaf nodes. Figure 3 illustrates the specific type of decision tree used in this paper. To describe this tree accurately, a graphical representation is required, showing one of the different types of decision trees used in previous studies as non-parametric regression tools to examine the impact of explanatory variables on the dependent variable [25]. Some researchers have adopted decision trees within the context of machine learning as a decision-making tool, employing established algorithms such as CART, GUIDE, and M5 (Alberto). In the present study, the GUIDE algorithm is employed to implement the decision tree framework, as illustrated in Figure 3 [2].

8.1 Generalized unbiased interaction detection and estimation (GUIDE)

It is one of the decision-making algorithms extensively studied by many researchers in the literature for building linear regression models as an important tool in the decision-making process [26]. This algorithm was designed to eliminate the bias in selecting the most significant variables in modeling important phenomena in regression analysis, thereby providing a good fit to the relevant experimental data [24]. Its operation is based on the chi-square test applied to residuals. The algorithm is constructed within the framework of piecewise constant linear regression models with univariate splits [27].

At each terminal node, the sample mean is calculated to serve as the estimate, followed by the computation of residuals. The node is then split into two groups: the first group contains positive residuals, while the second group contains non-positive residuals. The idea behind this division is to detect random patterns within these two groups using a sign test on the residuals at each node. A chi-square test can be employed to examine the association between the signed residuals at each node (categorized as belonging or not belonging) represented as rows in the test, and the total frequencies of the predicted values for c splits represented as columns [28]. If the predictive variable y is categorical with ccc categories, a $2 \times c$ contingency table is constructed, where the two rows correspond to the two residual groups and the c columns correspond to the predicted categories [29]. In the case where y is a quantitative variable, these values can be divided into columns according to a specific scheme, commonly using four groups representing the quartiles of the sample. This results in a 2×4 contingency table [30].

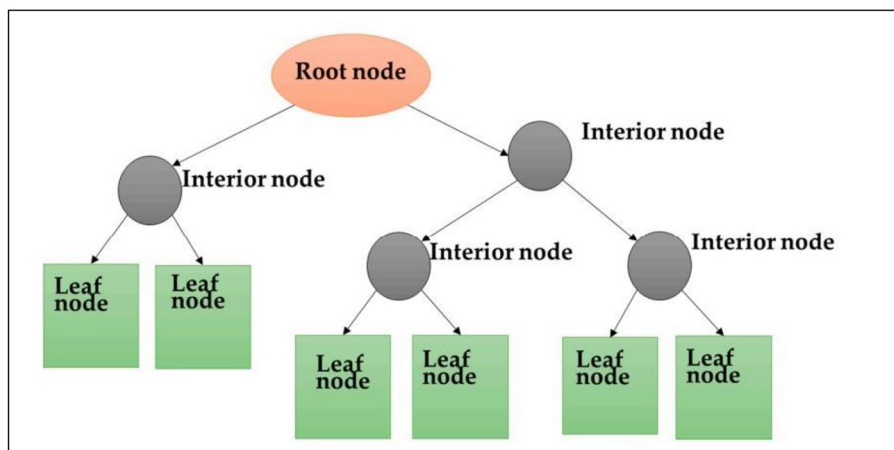


FIGURE 3 Decision tree nodes, root, interior and leaf [33].

8.2 Fuzzy decision tree regression structure

Binary decision trees are used to estimate the parameters of non-parametric regression models in the context of fuzzy set theory. Suppose that the function $\mu_A(x)$ represents the membership function of an element x to the fuzzy set A , and we attempt to find a solution by designing a decision tree with outcomes {either $x \in A$ or $x \ni A$ } where the value 1 represents membership of the element x to the set A , and the value 0 represents non-membership of the element x to the set A [23].

The degree of membership of the element x to the fuzzy set A allows for any real value between 0 and 1, which means the following [31]:

$$\mu_A(x) \in [0, 1].$$

By using partial membership in a fuzzy set, the strict logical constraints of set membership can be relaxed, thereby improving the performance of decision trees when dealing with fuzzy sets. This results in enhanced performance of decision trees in regression models in terms of flexibility and robustness [3].

Based on the literature of crisp decision trees, the GUIDE method can be used to estimate the regression model. Suppose there are P explanatory (independent) variables.

$$\text{Let } x_i = (x_1, x_2, \dots, x_p) \quad i = 1, \dots, p$$

Categorical variables do not pose a significant challenge, except in the case of fuzzy partitions of categorical variables. The values of the response variable represent the prediction target and can be either categorical (for classification) or real-valued (for regression) [32].

According to the definition of a fuzzy decision tree, all nodes of the tree are connected to the root node t_0 , where the decision tree is constructed using a hierarchical splitting strategy. The feature space is divided through a hierarchical sequence of logical tests into a set of non-overlapping regions, making the decision-making process straightforward. Each internal node in the decision tree corresponds to a test in the hierarchical structure used to construct the decision tree of interest [2].

Suppose the decision tree has been constructed up to a certain level, and the binary (terminal) nodes t_i are characterized by the membership function $\mu_i(x)$. Each node is split into two branches: one representing the value 1, which satisfies the logical test, and the other representing the value 0, which does not satisfy the logical test, as defined by Equation 5 [1].

$$\mu_{i\alpha}(x) = \mu_i(x) \mu_\alpha^{(i)}(x), \quad \alpha = L, R \tag{5}$$

The absolute membership degree and the number of training examples in the terminal node t_1 are expressed as Equation 6.

$$N_i = \sum_{n=1}^{N_{train}} \mu_i(x)_n \tag{6}$$

In a regression problem, the node t_i provides a prediction for the value of the response variable, which is equal to the mean of the response variable y for the training samples associated with t_i and it is represented as Equation 7 [33].

$$\bar{y}_i = \frac{1}{N_i} \sum_{n=1}^{N_{train}} \mu_i(x)_n y_n \tag{7}$$

In the regression model, the binary (terminal) node t_i represents the predicted value of the response variable. This value corresponds to the average of the response variable y calculated from the training samples assigned to node t_i . The mathematical formula for computing this average is given by Equation 8 [2].

$$\bar{y}_i = \frac{1}{N_i} \sum_{n=1}^{N_{train}} \mu_i(x)_n y_n \tag{8}$$

where

N_{train} Represents the number of training samples for each $n = 1, \dots, N_{train}$.

$\mu_i(\cdot)$ Represents the membership function for each $i = 0, 1, 2, \dots$, where i denotes the number of nodes in the tree.

$y_n \forall n = 1, \dots, N_{train}$ All elements of the response variable in the training set N_{train} .

$x_n \forall n = 1, \dots, N_{train}$ All elements of the dependent variable in the training set N_{train} .

Equation 2 can be expressed through building the tree as given in Equations 9, 10:

$$\bar{y}_1 = \frac{1}{N_1} \sum_{n=1}^{N_1} \mu_1(x)_n y_n \tag{9}$$

$$\bar{y}_2 = \frac{1}{N_2} \sum_{n=1}^{N_2} \mu_2(x)_n y_n \tag{10}$$

The terminal (leaf) nodes for all branches of the tree are calculated and denoted by \tilde{T} , such that the total number of nodes in the tree can be computed using the following formula [28].

$$2 \left| \tilde{T} \right| - 1.$$

Similarly, the number of internal (branch) nodes of the tree can be determined according to the following formula [26]:

$$\left| \tilde{T} \right| - 1.$$

The terminal nodes are used as predictive variables ($\bar{y}_1, \bar{y}_2, \dots$), and the average value of the dependent variable within each terminal node is calculated based on the number of observations that satisfy the test at that terminal node. Referring back to the definition in the formula above [31]. When applying the decision tree to a regression problem, it is essential to understand the membership relationship between the terminal (branch) nodes and the root node through the absolute degree of this membership, based on the above Equation 5 can be rewritten in a more detailed form as follows.

$$\mu_{i\alpha}(x) = \mu_i(x) \mu_{\alpha}^{(i)}(x), \quad \alpha = L, R. \tag{11}$$

Where R: right node, L: left node, and $\mu_i(x)$. The absolute degree of membership for the original node t_i can be calculated by repeatedly applying Equation 11 up to the root node. At this point, all points belong to the root node, Therefore, Equation 12 is holds true [24]:

$$\mu_0(x) = 1 \quad \forall x \tag{12}$$

All successful splits originating from the root node, repeated recursively, represent the sequence of connected logical tests from the root node to the terminal node t_i . This implies that the estimate (mean) at each terminal node corresponds to the outcome of the test at that terminal node. The sum of these estimates across all terminal nodes provides the overall estimate (global mean) of all decision nodes in the regression tree, which corresponds to the overall mean in a general regression model. Therefore, the

following Equation 13 must hold at every internal (splitting) node to ensure the consistency and accuracy of the model [33].

$$\mu_R^{(i)}(x) + \mu_L^{(i)}(x) = 1 \tag{13}$$

In the regression problem, the predicted value y given by the tree for a specific input feature vector x_{test} is expressed by the Equation 14:

$$\bar{y}_{(test)} = \sum_{t_l \in \tilde{T}} \mu_l(x_{test}) \bar{y}_l \tag{14}$$

Where l denotes the test set size (i.e., the number of observations in the test set).

\bar{y}_l is given in Equation 6 such that, by construction, only one of the membership values $\{\mu_l(x_{test}); t_l \in \tilde{T}\}$ equals 1, while the rest are zero. The error rate of the tree based on the training set is given by Equation 15 [1].

$$R_{train}(T) = \frac{1}{N_{train}} \sum_{n=1}^{N_{train}} (y_n - \bar{y}(x_n))^2 \tag{15}$$

Unless a stopping criterion is specified, the decision tree continues to grow until $R_{train}(T) = 0$:

However, there exists an optimal-sized tree beyond which, despite the fact that $R_{train}(T)$ continues to decrease monotonically, the predictive performance deteriorates (i.e., the true error rate or the unbiased estimation error increases). To avoid the problem of overfitting in the decision tree built from the training set and to obtain a tree of optimal size, several strategies have been proposed.

One of the most important methods to address excessive branching in tree growth is pruning, which involves removing branches that contain insignificant splits [5].

8.3 Chi-square tests with GUIDE algorithm

For the implementation of the work, the GUIDE algorithm processes four types of data (see Loh [30, p. 12]) as follows

- (1) *n*-variable: A numerical variable used for both estimation and node splitting.
 - (2) *F*-variable: A numerical variable used only for estimation and not for node splitting.
 - (3) *S*-variable: A numerical variable used only for node splitting and not for estimation.
 - (4) *C*-variable: A categorical variable used only for node splitting and not for estimation [1].
- (i) Extract the residuals resulting from fitting a constant model to the Y variable data.
 - (ii) For each numerical-valued variable, divide the data into four groups at the sample quartiles; construct a 2×4 contingency table with the groups as columns and the signs of the residuals (positive vs. non-positive) as rows; count the number of

observations in each cell and compute the χ^2 -statistic and its theoretical p -value from a χ^2 distribution.

- (iii) To detect interactions between a pair of numerical-valued variables (X_i, X_j), divide the (X_i, X_j)-space into four quadrants by splitting the range of each variable into two halves at the sample median; construct a 2×4 contingency table using the residual signs as rows and the quadrants as columns; compute the χ^2 - statistic and p -value. Again, columns with zero column totals are omitted. We refer to this as an interaction test.
- (iv) For each pair of variables (X_i, X_j), where X_i is numerical-valued and X_j is categorical, divide the X_i -space into two at the sample median and the X_j space into as many sets as the number of categories in its range [if X_j has c categories, this splits the (X_i, X_j), space into $2c$ subsets]; construct a $2 \times 2c$ contingency table with the subsets as columns and the signs of the residuals as rows; compute a χ^2 - statistic and p -value for the table after omitting columns with zero totals. If the smallest p -value is from a curvature test, it is natural to select the associated X variable to split the node. If the smallest p -value is from an in traction test, we need to select one of the two interacting variables. We could choose on the basis of the curvature p -values of the two variables but because the goal is to fit a constant model in each node, we base the choice on reduction in SSE [2]. This research relied on the fourth type of data (A categorical variable used only for node splitting and not for estimation).

9 Applied real data

One of the most prominent indicators of climate change is the rise in temperatures, a phenomenon that has had a clear impact in Iraq due to its geographical location between latitudes 29° and 37°N and longitudes 39° and 48°E . The country is characterized by a hot, semi-arid climate with generally moderate winds throughout the year [24].

Dhi Qar Governorate is located in southern Iraq, and as such, it is among the regions most affected by rising temperatures, which range between 14 and 51.1°C . Additionally, it experiences relatively higher humidity levels compared to other governorates. Meanwhile, wind speed is considered one of the naturally available

energy sources, and it has become increasingly important in recent years as it is classified as a clean energy source [34]. Moreover, it is regarded as one of the renewable energy resources, which are used for environmental preservation and in various fields, including electricity generation, sailing ships, and transportation, among others. In addition, it is affected by and influenced by the phenomenon of climate change occurring in the environment [33].

Notably, the highest average wind speed was recorded in 2007 at 3.1 m/s, while in 2025, the highest average reached 14.6 m/s. Data on temperatures and wind speed are obtained for a sample size of 425 daily observations in Dhi Qar Governorate for the period (December 2024–February 2025). Table 1 shows the descriptive statistics of the sample.

Table 1 illustrates that the average temperature variable was 32.4934°C , with the lowest recorded temperature for the studied period being 14°C and the highest being 51.1°C , which is relatively high. Regarding the wind speed variable, the average was 7.5256 m/s, and the highest recorded speed was 14.6 m/s. No wind speed was recorded for typical days, and this variable significantly impacts the climate in its minimum conditions.

9.1 Fuzzy logic description of the sample

In this paper, we presented a new statistical method for measuring and analyzing climate change in a specific geographical area and in a specific period of time, relying on data on two important variables (temperature and wind speed) for the purpose of identifying the risks of this phenomenon and taking appropriate precautions in the near and distant future to deal with this natural emergency that is increasing over time. The tool of this study the decision tree regression method, was employed as a new extension of its application in the phenomenon of climate change, relying on two variables and using fuzzy logic to divide the data into groups as an accurate description of the climate situation.

Here we must study and analyze the emerging phenomena according to the reference literature for those interested in studying the field of knowledge from all its aspects and relationships. Here, we employed the decision tree approach as a primary tool, relying on it to construct a regression model for the fuzzy data of both study

TABLE 1 Descriptive statistic for that actual data of the two variables (temperature and wind speed).

Variables	No.	Mean	Min	Max	First quarter (q1)	Second quarter (q2)	Third quarter (q3)
Temperature (X_1)	425	32.4934	14	51.1	22.3	29.5	43.7
Wind speed(X_2)	425	7.5256	0	14.6	5.8	7.4	9.15

TABLE 2 Represented attributes vs. fuzzy sets for the variables (temperature and wind speed).

Variables	A: Fuzzy sets					
	attributes	set1	set2	set3	set4	Set5
Temperature (X_1)	(Language variables in Fuzzy Logic)	Very cold	Cold	Mild	Hot	Very hot
Wind speed (X_2)		Very slow	Slow	Mild	Strong	Very strong

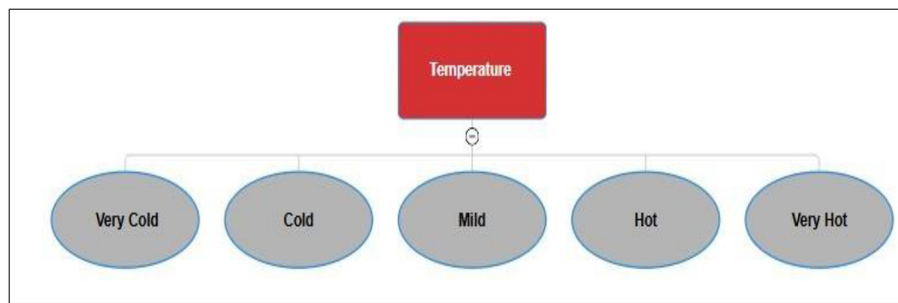


FIGURE 4
The Attributes of the Temperature Variable (very cold, cold, mild, hot, very hot).

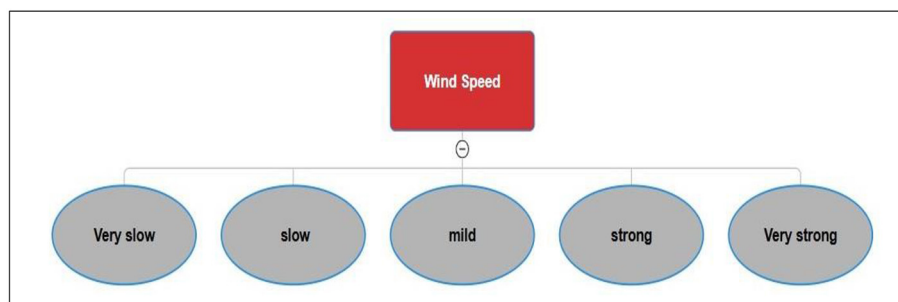


FIGURE 5
The Attributes of the wind speed Variable (very slow, slow, mild, strong, very strong).

variables. In the previous section, we presented the quantitative and fuzzy general description of the two variables (Tables 1, 2).

We utilized fuzzy logic to transform the quantitative and real-valued data of the two variables temperature, and wind speed, which are characterized by ambiguity and uncertainty, into categorical variables represented by clearer attributes and partitions. These transformed variables may simultaneously belong to two groups at the same time, as shown in Table 2. Figure 4 illustrates the fuzzy membership functions (triangular) for each partition (attribute) of the temperature variable, as described in Table 2.

9.2 Fuzzy regression tree description

In this section, in line with the objectives of the research, the two variables are examined separately and in a detailed manner. Figures 4, 5 present the partitions of each variable based on their respective root nodes adopted in this study. Furthermore, Figures 6a–e, 7a–e provide a comprehensive illustration of the hierarchical branching structure at each decision node of the decision tree model constructed using the real data corresponding to the two variables, respectively. In light of this approach, the binary-division sets (binary-division decision trees type) for the purpose of employing the chi-square test, the data set is divided into two types: first, the negative-signal set and second, the positive-signal set. Then, each node of the tree is constructed to represent

a specific fuzzy partition based on fuzzy (linguistic) rules. This process transforms the crisp input data into fuzzy sets relying on membership functions, which in turn act as splitting nodes in the tree branching into five child nodes. According to the hierarchical structure of the decision tree partitions, these form partial groups based on quartiles. Subsequently, these groups are further divided into two terminal (leaf) nodes corresponding to positive and negative difference signals. Based on the calculated test statistic, decisions are made using the contingency tables method for the chi-square test. The related tables and the following section explain in detail.

To carry out the analysis using the steps of the GUIDE algorithm discussed in Section 8.1 on the data of the climate change variables (temperature and wind speed), we will construct 2×4 contingency tables for the chi-square test to measure the independence of the error sign between the predicted and actual values. This will be done for both the observed and expected frequencies across the four fuzzy group partitions (columns). Then, we calculate the chi-square test statistic and compare it with the tabulated critical value ($\chi^2_{(3,0.05)} = 7.8147$) according to the following hypothesis testing: the null hypothesis, which represents the levels of sign error that are independent of the levels for the fuzzy partition groups, against the alternative hypothesis, which represents the levels of sign error that are dependent on the levels for the fuzzy partition groups. Figures 8a, b explained the descriptive for fuzzy partition groups.

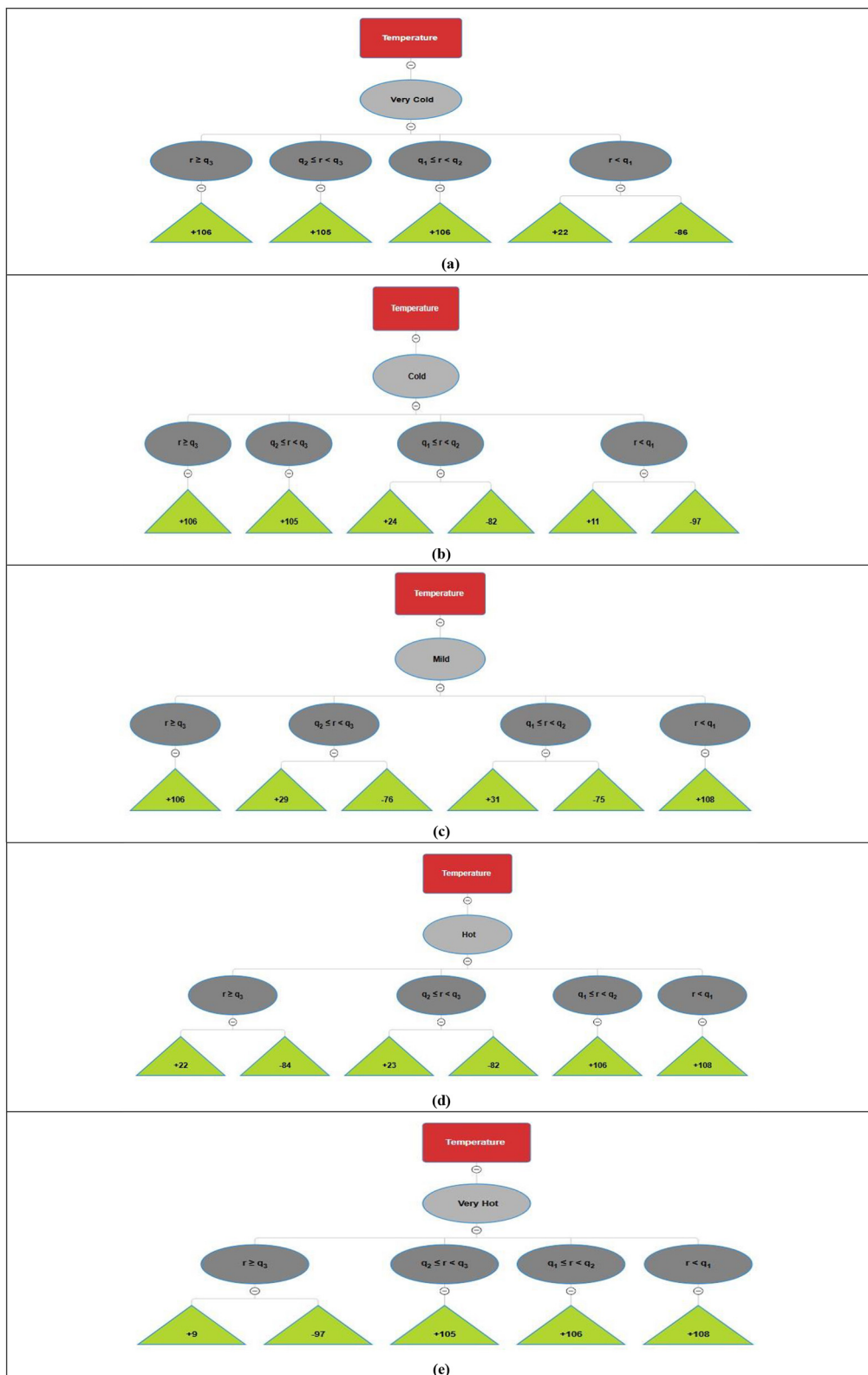


FIGURE 6 The Fuzzy Regression Tree Models for the Attributes of the actual data set representing Variables Temperature (a) Internal nodes of the very cold category, (b) Internal nodes of the cold category, (c) Internal nodes of the mild category, (d) Internal nodes of the hot category, (e) Internal nodes of the very hot category.

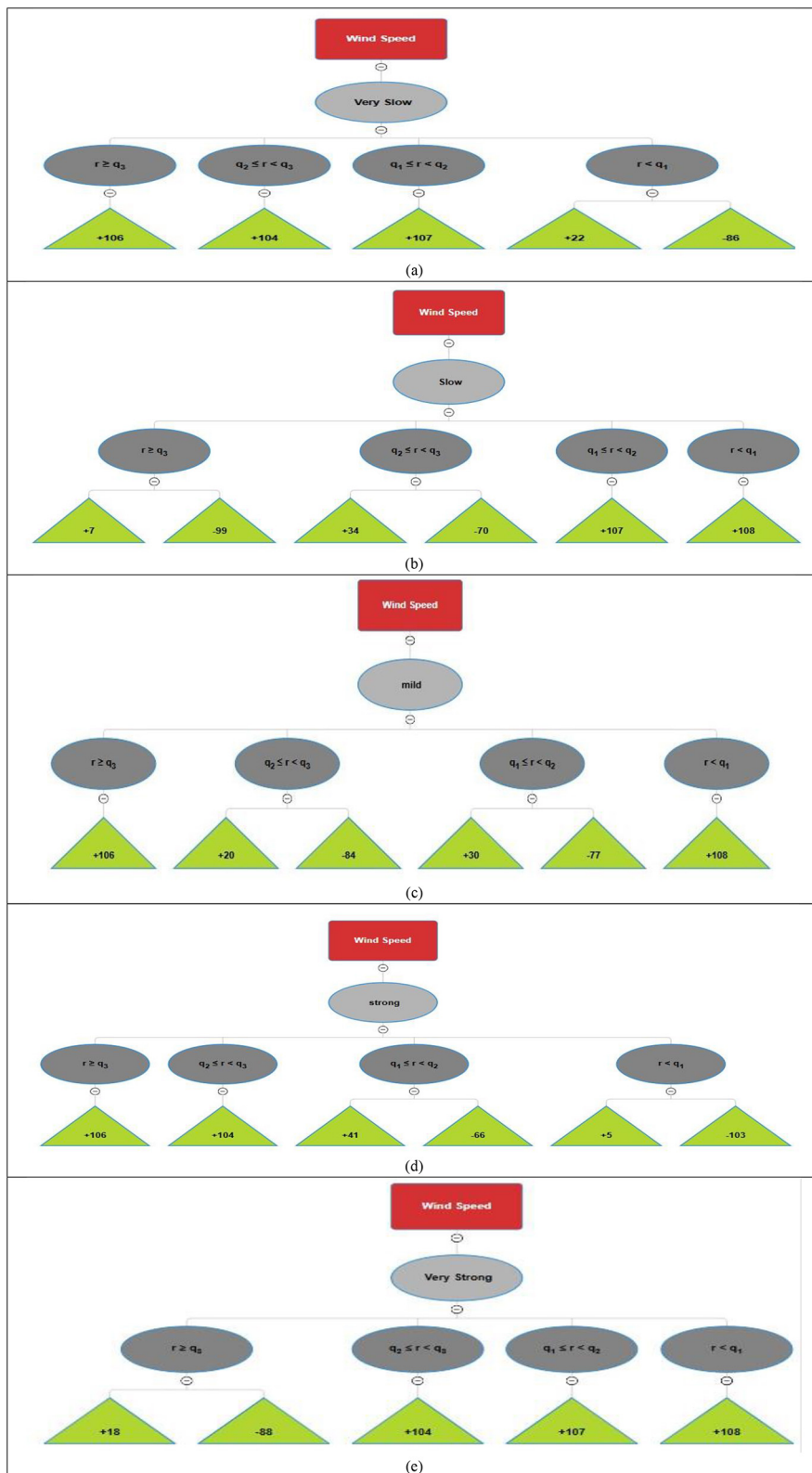


FIGURE 7 The Fuzzy Regression Tree Models for the Attributes of the actual data set representing Variables wind speed (a) Internal nodes of the very slow category, (b) Internal nodes of the slow category, (c) Internal nodes of the mild category, (d) Internal nodes of the strong category, (e) Internal nodes of the very strong category.

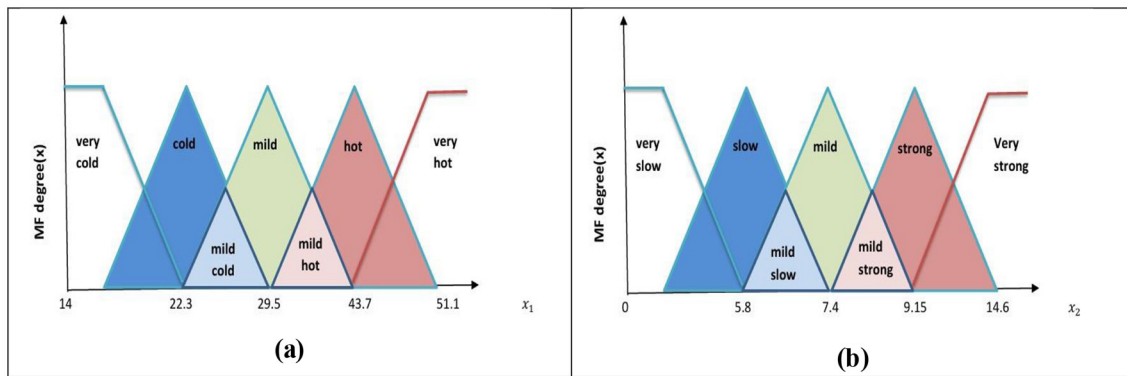


FIGURE 8 Triangular membership function for the attributes of the actual data set representing: (a) Variables Temperature (very cold, cold, mild, hot, very hot), (b) Variables wind speed (very slow, slow, mild, strong, very strong).

TABLE 3 Observation count and expected count of the fuzzy number for temperature.

	Temperature	Sign error	$r < q_1$	$q_1 \leq r < q_2$	$q_2 \leq r < q_3$	$r \geq q_3$	\sum Count = 425
Observation count	Very Cold	+	22	106	105	106	339
		-	86	0	0	0	86
Expected count	Very Cold	+	86.146	84.551	83.753	83.753	-
		-	21.854	21.450	21.247	21.247	-
Observation count	Cold	+	11	24	105	106	246
		-	97	82	0	0	179
Expected count	Cold	+	62.513	61.355	60.776	61.355	-
		-	45.487	44.645	44.124	44.645	-
Observation count	Mild	+	108	31	29	106	274
		-	0	75	76	0	151
Expected count	Mild	+	69.628	68.339	67.694	68.339	-
		-	38.372	37.661	37.306	37.661	-
Observation count	Hot	+	108	106	23	22	259
		-	0	0	82	84	166
Expected count	Hot	+	65.816	64.598	63.988	64.598	-
		-	42.184	41.402	41.012	41.402	-
Observation count	Very Hot	+	108	106	105	9	328
		-	0	0	0	97	97
Expected count	Very Hot	+	83.351	81.807	81.053	81.807	-
		-	24.649	24.193	23.965	24.193	-

Figure 8a illustrates the triangular membership functions for the temperature variable. And Figure 8b illustrates the triangular membership functions for the wind speed variable, which are measuring some climate change factors during the study period in Dhi Qar Governorate. and the characteristics that represent the fuzzy sets used to classify the temperature variable into five fuzzy groups (very cold, cold, mild, hot, and very hot) and the wind speed variable into five fuzzy groups (very slow, slow, mild, strong, and very strong). In order to give a clear and accurate perception of

the weather condition and remove the ambiguity from the general description of the condition to a more accurate description closer to measuring.

Figure 4 showed the definition of the root node of the temperature variable, and Figure 5 showed the definition of the root node of the wind speed variable. Figures 6a–e shows the branches of the regression tree of the temperature variable for the fuzzy partitions that we explained in Figure 8a and which relied in its calculations on the steps of the GUIDE algorithm to clarify the

TABLE 4 Observation count and expected count of the fuzzy number for wind speed.

	wind speed	Sign error	$r < q_1$	$q_1 \leq r < q_2$	$q_2 \leq r < q_3$	$r \geq q_3$	\sum Count = 425
Observation count	Very Slow	+	22	107	104	106	339
		-	86	0	0	0	86
Expected count	Very Slow	+	86.146	85.348	82.955	83.753	-
		-	21.854	21.652	21.045	21.247	-
Observation count	Slow	+	108	107	34	7	256
		-	0	0	70	99	169
Expected count	Slow	+	65.054	64.452	62.645	63.849	-
		-	42.946	42.548	41.355	42.151	-
Observation count	Mild	+	108	30	20	106	264
		-	0	77	84	0	161
Expected count	Mild	+	69.628	66.466	64.602	65.845	-
		-	38.372	40.534	39.398	40.155	-
Observation count	Strong	+	5	41	104	106	256
		-	103	66	0	0	169
Expected count	Strong	+	65.054	64.452	62.645	63.849	-
		-	42.976	42.548	41.355	42.151	-
Observation count	Very Strong	+	108	107	104	18	337
		-	0	0	0	88	88
Expected count	Very Strong	+	85.638	84.845	82.466	84.052	-
		-	22.362	22.155	21.534	21.948	-

work using this approach and relying on the analysis in Table 3. For the purpose of clarifying the work in Table 3, (r) represents the number of repetitions in each cell that corresponds to the error signal for each of (the observed repetition and the expected repetition) and the corresponding subset of the total sets that were divided through the three quartile values ($q_1 = 22.3, q_2 = 29.5, q_3 = 43.7$). Figures 7a–e shows the branches of the regression tree for the wind speed variable for the fuzzy partitions that we explained in Figure 8b and which relied in its calculations on the steps of the GUIDE algorithm to clarify the work using this approach and relying on the analysis in Table 4. For the purpose of clarifying the work in Table 4, (r) represents the number of repetitions in each cell that corresponds to the error signal for each of (the observed repetition and the expected repetition) and the corresponding subset of the total sets that were divided through the three quartile values ($q_1 = 5.8, q_2 = 7.4, q_3 = 9.15$) in fuzzy logic.

Tables 3, 4 display the classification outcomes derived from the analysis of a fuzzy dataset. These results were validated using two independent tree-based models. Figures 6a–e, 7a–e demonstrate the underlying decision-making process, which leverages the inherent tree structure within fuzzy regression. This process involves the division of the sample space into quartiles to provide a comprehensive description of the variable states. In the case of analyzing the temperature variable, the output results are in Table 3, where the second and third columns display the fuzzy group divisions and the error signal for each division. The rest of the columns represent the branching results for the variable

according to the fuzzy categorical division of the error signal. Thus, the descriptive results of the repetition of cases can be analyzed in the cells corresponding to the chi-square table as follows: As for the first division, very cold weather has a main influence and is the focus of our attention, where 339 observations with a rate of 80% corresponded to descriptions of cold weather (the error sign is positive), which compares with rate 20% for very cold weather. Analyzing the first quartile, 22 observations with a rate of 20% were classified as cold weather, while 86 observations with rate 80% were categorized as very cold weather. As for the remaining quarters of this category division, they were 106 and 105 observations, with rates of 100% for cold and rates of 0% for very cold, respectively. As for the rest of the category classifications, they can take the same analysis, and they are moderate with mild climate category divisions and are considered not influential in the main way (cold, intermediate, and hot). As for the fifth division, very hot, it has a main influence and is the focus of our attention. We recorded 328 observations, with 77% of the total observations, compared to 23% who described the weather as very hot. As for the fourth quarter, which described the weather as hot, there were 9 observations, with a rate of 8.5%, while 97 observations described the weather as very hot, representing 91.5%. For Table 3, we found that the calculated chi-square value for the very cold of the temperature variable was 316.4632, and compared with the corresponding tabular chi-square value, which was $[\chi^2_{(3,0.05)} = 7.8147]$, we note that the calculated value is greater than the tabular value, and thus the null hypothesis is rejected and the hypothesis that states that the levels of the

residuals signal are not independent of the partial sums is accepted, and this means that the first fuzzy set for the temperature variable affects climate change. And, we found that the calculated chi-square value for the very hot of the temperature variable was 378.2436, and compared with the corresponding tabular chi-square value, which was $[x^2_{(3,0.05)} = 7.8147]$, we note that the calculated value is greater than the tabular value, and thus the null hypothesis is rejected and the hypothesis that states that the levels of the residuals signal are not independent of the partial sums is accepted, and this means that the Fifth fuzzy set for the temperature variable affects climate change.

We conclude from Table 3 that the fuzzy classifications were more accurate in describing temperatures when using the fuzzy regression tree structure based on the selected categorical classifications provided by fuzzy logic. This improves the accuracy of the results shown in Table 3 and provides a clear understanding for both the general public in Dhi-Qar Governorate who are interested in weather conditions and researchers studying climate change, enabling decision-makers to take immediate and future action to mitigate and prevent temperature increases and to utilize them in electricity generation by developing alternative plans and environmentally friendly policies.

When analyzing the results for the wind speed variable in Table 4, the second and third columns display the fuzzy group classifications and the error signal for each classification. The remaining columns present the results of the variable's branches according to the fuzzy categorical division of the error signal. Accordingly, the descriptive results of case repetitions can be analyzed in the cells corresponding to the chi-square table as follows: The number of observed frequencies in the sample describing the weather as strong winds was 256 observations, accounting for 60%, compared to 40% who described the weather as stormy. In the first and second quarters, 46 observations, a rate of 21% described the weather as having strong winds, while 169 observations, a rate of 79%, described it as stormy. The number of observations in the sample describing the weather as stormy was 337, accounting for 79%, compared to 21% who described it as very stormy. In the fourth quarter, 18 observations, a rate of 17% described the weather as stormy, while 88 observations, a rate of 83% described it as very stormy. For Table 4, we found that the calculated chi-square value for the strong of the wind speed variable was 302.1632 and compared with the corresponding tabular chi-square value, which was $[x^2_{(3,0.05)} = 7.8147]$, we note that the calculated value is greater than the tabular value, and thus the null hypothesis is rejected and the hypothesis that states that the levels of the residuals signal are not independent of the partial sums is accepted, and this means that the fourth fuzzy set for the wind speed variable affects climate change. And, we found that the calculated chi-square value for the very strong wind speed variable was 333.9847 and compared with the corresponding tabular chi-square value, which was $[x^2_{(3,0.05)} = 7.8147]$, we note that the calculated value is greater than the tabular value, and thus the null hypothesis is rejected and the hypothesis that states that the levels of the residuals signal are not independent of the partial sums is accepted, and this means that the fifth fuzzy set for the wind speed variable affects climate change.

We conclude from Table 4 that the structure of the fuzzy regression tree, based on the selected categorical classifications provided by fuzzy logic, played a significant role in monitoring weather. Analyzing the results presented in Table 4 provides information both the general public in Dhi-Qar Governorate who are interested in weather, and for researchers focused on climate change and alternative energy. These findings can assist decision-makers in utilizing wind speed as a source of clean and renewable energy. Wind energy can be harnessed across various fields, with one of the most important being the generation of electricity in an environmentally friendly manner.

10 Conclusions

The GUIDE decision tree algorithm has been widely applied in the literature to analyze crisp datasets. The main contribution of the present study is the extension of its application to the framework of fuzzy logic theory, which clearly differentiates this work from existing studies. We conclude that fuzzy classifications were more accurate in describing temperature and wind speed variables when using the fuzzy regression tree structure based on the selected categorical classifications provided by fuzzy logic. In addition, these variables played a significant role in monitoring weather conditions. In other words, fuzzy logic offers a more accurate description and provides a clearer picture of the significant and rapid changes in climate and their impact on the environment. It has become evident that regression tree tools are among the most efficient and accurate methods for quantitative analysis, aiding in making correct and precise decisions. There is clear evidence of climate change caused by the combined influence of rising temperatures and wind speed in the southern regions of Iraq, particularly in Dhi-Qar Governorate.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

AH: Writing – original draft, Writing – review & editing. HH: Writing – original draft, Writing – review & editing. AA-S: Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Suárez A, Lutsko JF. Globally optimal fuzzy decision trees for classification and regression. *IEEE Trans Pattern Anal Mach Intell.* (1999) 21:1297–311. doi: 10.1109/34.817409
- Loh WY. Classification and regression trees. *Wiley Interdiscip Rev Data Mining Knowl Discov.* (2011) 1:14–23. doi: 10.1002/widm.8
- Segatori A, Marcelloni F, Pedrycz W. On distributed fuzzy decision trees for big data. *IEEE Transac Fuzzy Syst.* (2017) 26:174–92. doi: 10.1109/TFUZZ.2016.2646746
- Mohammadiun S, Hu G, Gharahbagh AA, Mirshahi R, Li J, Hewage K, et al. Optimization of integrated fuzzy decision tree and regression models for selection of oil spill response method in the Arctic. *Knowl Based Syst.* (2021) 213:106676. doi: 10.1016/j.knsys.2020.106676
- Nikolaidis P. Wind power forecasting in distribution networks using non-parametric models and regression trees. *Discover Energy.* (2022) 2:6. doi: 10.1007/s43937-022-00011-z
- Zimmermann HJ. *Fuzzy Set Theory—and Its Applications.* Berlin: Springer Science and Business Media (2011).
- Habeeb AS. Estimating the parameters of the odd Lomax exponential distribution. *Stat Optim Inform Comput.* (2025) 13:694–715. doi: 10.19139/soic-2310-5070-2121
- Spolaor S. *Fuzzy Logic for the Modeling and Simulation of Complex Systems.* Milan: University of Milano-Bicocca (2020).
- Buckley JJ, Eslami E. *An Introduction to Fuzzy Logic and Fuzzy Sets, Vol. 13.* Berlin: Springer Science and Business Media. (2002). doi: 10.1007/978-3-7908-1799-7
- Hasan HA, Mohammad MJ. Classify the nutritional status of Iraqi children under five years using fuzzy classification. *Sumer J. Pure Sci.* (2024). 29, 161–171. doi: 10.33095/jeas.v29i138.3046
- Hellmann M. *Fuzzy Logic Introduction.* Rennes: Université de Rennes (2001), 1.
- Zhang H, Liu D. *Fuzzy Modeling and Fuzzy Control.* Berlin: Springer Science and Business Media (2006).
- Bojadziev G, Bojadziev M. *Fuzzy Logic for Business, Finance, and Management (2nd Ed.).* Singapore: World Scientific Publishing Co. Pte. Ltd. (2007).
- Hooda DS, Raich V. *Fuzzy Logic Models and Fuzzy Control: An Introduction.* Oxford: Alpha Science International Ltd. (2017).
- Hasan HA, Mohammad MJ. Classification of Iraqi children according to their nutritional status using fuzzy logic. *J. Econ. Administr. Sci.* (2023) 29:161–71. doi: 10.33095/jeas.v29i138.3046
- Ross TJ. *Fuzzy Logic with Engineering Applications.* Hoboken, NJ: John Wiley and Sons (2009).
- Suganthi L, Iniyan S, Samuel AA. Applications of fuzzy logic in renewable energy systems—a review. *Renew Sustain Energy Rev.* (2015) 48:585–607. doi: 10.1016/j.rser.2015.04.037
- Zimmermann HJ, Zadeh LA, Gaines BR. Fuzzy sets decision analysis. *Fuzzy Sets Syst.* (1985) 1:45–65. doi: 10.1016/0165-0114(78)90031-3
- Mann PS. *Introductory Statistics.* Hoboken, NJ: John Wiley and Sons (2010).
- Alavala CR. *Fuzzy Logic and Neural Networks: Basic Concepts and Applications.* Bengaluru: New Age International Publisher (2008).
- Nowaková J, Pokorný M. Fuzzy linear regression analysis. *IFAC Proc.* (2013) 46:245–9. doi: 10.3182/20130925-3-CZ-3023.00079
- Shapiro AF. Fuzzy regression models. *J Optimiz Theor Appl.* (2005) 102:373–83. doi: 10.1023/A:1021706631165
- Hasan HA, Mohammad MJ. Classification of Iraqi children according to nutritional status using fuzzy decision tree. *J Al-Rafidain Univ Coll Sci.* (2025) 56:468–80. doi: 10.55562/jruc.v56i1.42
- Magesh T, Thiyagesan M. Machine learning-driven wind energy forecasting for sustainable development. *MATEC Web Conf.* 393:02003 (2024). doi: 10.1051/mateconf/202439302003
- Marsala C. Fuzzy decision trees for dynamic data. In: *2013 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)* (2013), 17–24.
- Yuanyuan Z. MOOC teaching model of basic education based on fuzzy decision tree algorithm. *Comput Intell Neurosci.* (2022) 2022:3175028. doi: 10.1155/2022/3175028
- Loh WY. Improving the precision of classification trees. *Ann Appl Stat.* (2009) 3:1710–37. doi: 10.1214/09-AOAS260
- Yu H, Lu J, Zhang G. Learning a fuzzy decision tree from uncertain data. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (2017), p. 1–7.
- Park Y. A comparison of neural net classifiers and linear tree classifiers: their similarities and differences. *Pattern Recognit.* (1994) 27:1493–503. doi: 10.1016/0031-3203(94)90127-9
- Loh WY. Regression trees with unbiased variable selection and interaction detection. *Stat Sin.* (2002) 12:361–86. <https://www.jstor.org/stable/24306967>
- Fathima TH, Kovoov BC, Ku J. Big data classification based on distributed fuzzy decision trees. In: *Proceedings of ICAEEC-2019, IIIT Allahabad India, 31st May-1st June, 2019* (2019). doi: 10.2139/ssrn.3576492
- Chen CH, Hårdle W, Unwin A, Loh WY. Regression by parts: fitting visually interpretable models with GUIDE. In: *Handbook of Data Visualization.* Berlin: Springer (2008). p. 447–69.
- Kassim NM, Santhiran S, Alkahtani AA, Islam MA, Tiong SK, Mohd Yusof MY, et al. (2023). An adaptive decision tree regression modeling for the output power of large-scale solar (LSS) farm forecasting. *Sustainability* 15:13521. doi: 10.3390/su151813521
- Yürek Ö, Birant D, Yürek I. Wind power generation prediction using machine learning algorithms. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi.* (2021) 23:107–19. doi: 10.21205/deuofd.2021236709
- Ross S. *Probability and Statistics for Engineers and Scientists.* New Delhi: Elsevier (2010).