

Comparison Fuzzy Time Series -Clustering Applications in Production and Consumption Electric Prediction

Mohammed Eid Awad Alqatqat*, Ma Tie Feng

Department of Statistics, Southwestern University of Finance and Economics, Chengdu, China

Abstract This study aim to Comparison in fuzzy time series and fuzzy time and clustering and Proposed Method in the field production and consumption electric, the basic idea of three methods is to predictive the Possibility of future production and consumption electric based on historical data of production and consumption electric in the past, Nevertheless, Three methods have a different approach in transforming the value of production and consumption to the range of the random variables (the states). This paper considered the production and consumption data of electric in china from January, 2015 to December, 2019. The accuracy of three methods is verified by using the Mean Absolute Deviation (MAD), Mean Absolute Percentage Error (MAPE), the result shows that the Proposed Method has smaller MAPE, MAD, From fuzzy time series and fuzzy time clustering methods.

Keywords Fuzzy time series, Clustering algorithm, Time Series, Prediction, Electric Energy

1. Introduction

Predicting the size of any phenomenon in the future is an important matter that helps in understanding the behavior of the phenomenon with time, and thus how to confront it. It is not possible to make future plans to confront the phenomenon except with its future dimensions and knowing the shape of these dimensions and their patterns. The consumption of electrical energy and its production within society has a dynamic character because it is not fixed, as it is affected by many variables, including the development of life and the increase in the necessary electrical appliances used in a necessary manner. So the scale of the problem is that electricity consumption and production now will not be the same after twenty years, for example.

To face this problem, it is necessary to make estimates that monitor its size in the coming years, as it is not possible to build long-term plans to address this problem without making quantitative expectations for the size of the problem.

The historical pattern of the problem can only be determined by the quantities that express its variables, as well as the increasing demand for electrical energy as a result of the development taking place in all facilities of life (industrial, educational, health and agricultural).

The idea of this paper aims at how to build and fit a model to find the best model in a way that can accurately predict the rates of consumption and production of electricity for the

coming years.

Technological advancement, population growth and increased economic activities have led to a rise of various issues in the planning of energy such as investment, allocation as well as search for a new sustainable source of energy [1]. With electricity being the most formidable and reliable source of energy at the moment, more advanced use for it as well as its increased development and advancement is in consideration [2]. As a way of resolving these rising issues, there is an importance of having a central unit under which all this advancements are monitored, controlled as well as optimized to ensure that the generation and transmission of energy is uniform and most of all safe. Such control should be provided for by the power generating companies or the government. Such a system can help in automatic obtaining of the load data from networks of retail stores, factories, as well as office buildings which can then be self-diagnosed after which a trend analysis can be conducted to help perform a forecast on such data accordingly [3]. Such forecasting can be of use to the electric utility to help in the making of vital decisions such as decision to purchase or generate power, development of infrastructure as well as switching the load. Forecasting can be described as an approach that's predictive analytical. Such an approach helps to deal with the prediction of the future through the use of past data as well as models [4]. The approach can also be used in different domains such as personal management, finance management, resource management and organizational management.

Globally, the study of electric load forecasting management is still ongoing [5]. Most of the entities more concerned with such research are the electricity producing

* Corresponding author:

centerload2012@gmail.com (Mohammed Eid Awad Alqatqat)

Received: July 30, 2020; Accepted: August 15, 2020; Published: September 5, 2020

Published online at <http://journal.sapub.org/statistics>

companies. Such studies help in making and determining how energy is going to be managed and planned by these firms. However, there are some methods that are already in place being used to forecast electricity. They include regression time series analysis, time series, autoregressive integrated moving average (ARIMA), artificial neural network (ANN), particle swarm optimization (PSO), Cluster analysis and the fuzzy time series (FTS) [6]. The cluster analysis and the FTS are the commonly used methods in forecasting since they can easily identify structures within data [7]. FTS uses different models which have been proposed by researchers to help in resolving problems. Some of the fields which use FTS include the enrolment of universities, stock prices index, financial sectors, evaluation of temperature and also electricity load consumptions [8]. The models used in FTS helps in the production of rules, systems and procedures used in forecasting which are all based on factors such as interval length of discourse universe, the weight of fuzzy relationship as well as the mathematical models [9]. These factors help in the determination of in-out sample forecasts. In order to achieve high levels of forecasting, the three factors are crucial for FTS [10].

1.1. Statement of the Problem

The problem is the increase in consumption of electrical energy over time, therefore it is necessary to know the variables that affect the consumption of electrical energy and then study the possibility of finding projections of demand for electrical energy.

1.2. Objectives of the Study

1. Learn about the trend of electricity production and consumption trend.
2. The importance of the role of the forecasting process in rationalizing decisions and exploring possible Consequences, defining the time series method and its effectiveness in the long-term forecasting process.
3. Building mathematical models that can predict the quantities of electricity consumed and produced per month based on accuracy criteria.

2. Summary of Literature Review

2.1. Review of Fuzzy Time Series

The regular set (crisp) is defined as a set of elements that is an item that can belong to or belong to a group and that the group may be specified or not defined [11], [12].

Whereas, the set fuzzy group [13], Is that it is a class of elements with an organic degree and that this group is distinguished by an organic function that allocates each An element of a membership degree whose range is between zero and one, i.e. when the element takes a membership degree (1), this means that the element belongs entirely to the fuzzy group and when the membership degree is (0), this means that the element belongs absolutely to the group and other degrees vary between zero and one, when The degree

of membership is (0.5), this means that the element belongs to the ratio (0.5) to the fuzzy group and does not belong to the group with the same percentage. The element belongs to the fuzzy group and does not belong to it with a ratio of (0.1) and this is closer to membership than to whether or not.

Other researchers have provided many definitions of the fuzzy group, as Kaufmann knew it (1975) as follows [14], The fuzzy group is the group in which there are precisely clear boundaries between those elements that belong to it and those to which they belong. As for the definition that he provided in 1988 by Zimmerman [15] then the most accurate definitions are as follows: If X is a group of elements symbolized by generally denoted by x , the A group in $A \times X$ is a set of arranged pairs, Is the membership function x in A which is a function from X to μ as $\mu \in A$ where (x) the continuous membership field in the closed period.

Fuzzy set has 2 attributes, namely:

1. Linguistics is the naming of a group that represents a situation or certain conditions using natural language, such as: cold, cool, normal, warm and hot.
2. Numerical is a value or number that indicates the size of a variable such as: 40, 25.50 and so on.

There are several things that need to be known in understanding fuzzy systems, namely [16]:

1. Fuzzy variables are variables that are to be discussed in a system fuzzy. Example: age, temperature, sales, demand and so on.
2. Fuzzy set is a group that represents a certain condition or condition in a fuzzy variable. Example: Age variable is divided into 3 fuzzy sets, namely YOUNG, PAROBAYA and OLD. Variable the temperature is divided into 5 fuzzy sets namely COLD, COOL, NORMAL, And WARM and HEAT.
3. Universe of Talk The universe of speech is the whole value that is allowed to be operated in a fuzzy variable. The universe of speech is a set of real numbers that always rise monotonically from left to right. Talking universe values can be positive and negative numbers.
4. The fuzzy set domain is the whole value that is allowed in the universe of speech and may be operated in a fuzzy set. Like the universe of speech, the domain is a set of real numbers that always rises monotonically from left to right. Domain values can be positive or negative numbers.

Literature done by Silva and Lisboa indicates that FTS model was introduced about two decades ago [10]. Since then, FTS has been used widely due to its

Superiority while dealing with knowledge that is imprecise such as linguistics in regard to decision making [10]. Literature also indicates that there is need for the improvement of FTS forecasting which has been done through proposing new methods [17]. Since its discovery, FTS approach has undergone some modification leading to improvement to near perfection such as Song's approach which offered a simpler forecast method in 1996 [18]. Some of the main issues with the FTS in regard to forecasting is the

interval length which greatly affects the accuracy of the model [9]. Some of literature has addressed the problem through adjusting the interval lengths using an optimizing or distribution technique [19].

Other researchers have focused on improving the weighted forecast models as a way of improving the accuracy of forecast results.

This model is responsible for chronological order and various recurrences [20]. Further literature indicates that there are many forecasting models which are original and based on FTS, which are presented as well as combined with novel algorithms and technologies. Some of such models according to Chen include a model proposed by Singh who proposed that FTS can be used to forecast methods of crop production based on different parameters of computational methods [4]. Research done by Lee and Hong indicates that several models of FTS such as genetic algorithm, the simulated annealing algorithm can help forecast temperature [2].

Research done by Ismail, Efendi and Deris indicates that the energy system in our case electricity has two aspect that is, supply and demand, since it is considered to be a product that has value [3]. Before 1960, the management of energy was concerned with prior aspect while the latter considered being a given data hence promoting the narrative of arranging for adequate supply to ensure that demand is met and well satisfied [11]. In the 1970s however, the prices hiked which led to researchers, governments as well as utilities to begin an investigation to the problem which was more focused on demand [21]. Ismail, Efendi and Deris defined energy demand management as “a systematic utility and government activities designed to change the amount and/or timing of customer’s use of energy” [11]. These activities also according to the literature involve utilizing the energy resources effectively, ensuring reliable supply, ensuring that the energy resources are managed efficiently, conservation of energy, a combination of power and heat systems, renewable energy systems, energy systems that are integrated and more so power delivery systems that are independent [22]. Research done by Hong and Lee indicated that the management of demand constitutes planning, implementation as well as monitoring of energy utilization all of which are designed to ensure that the consumers are encouraged to modify their pattern and level of energy usage [11]. Chen and Hong indicated that, in the prediction of energy consumption and demand, fuzzy time series is the best model since it bases the prediction on the size of the population, the methods used as well as the forecasting horizons [4]. Literature by Yu also indicates that the FTS has been used to forecast energy consumption especially the forecasting regional electricity loads [23]. FTS is also used help in the reduction of load forecasting error in short-term load forecasting problems [24]. Electricity producers and suppliers also use the FTS to estimate and predict the demand of electricity for monthly and seasonal changes.

Research done by Ozawa and Niimura indicated that FTS has been used in countries such as Turkey to forecast

short-term gross annual electricity demand by simply taking into consideration the political and economic as well as the electricity market conditions [5]. These are the same factors that are considered for any form of good which follows the rules of demand and supply. The two researchers also indicated that FTS is usually used or rather implemented while conducting research of fuzzy neural network for monthly demand of electricity forecasting as well as while using neural networks which helps in analyzing of several approaches while predicting the consumption of natural gas in different regions [25]. Countries such as Poland and Taiwan have gone on record as users of these models while using FTS. The FTS concept according to literature is a new approach which was developed through resolving problems of linguistic time series data [26]. FTS is an approach that uses a combination of linguistic variables and the analysis process of applying fuzzy logic into time series to help in solving data fuzziness [27]. Literature by Bolturk, Oztaysi and Sari indicated that FTS has an important aspect which is the assumption in regard to data which is not needed [6]. This assumption helps to differentiate FTS from other statistical approaches.

2.2. Review of Cluster Analysis

After the person discovered that there are common characteristics in the parts of knowledge about me by classifying them, and philosophers and thinkers made efforts to lay the foundations and systems for this classification, and thus the word "classification" in the language is to distinguish things from each other in a way arranged in classes or sections and if arranged and known, it has been prepared A kind of classification system. Accordingly, classification can be defined as the process of grouping similar things together and all members of the first group, department, or single class resulting from classification participate in at least one specific characteristic that members of departments or other classes do not have to benefit from in scientific fields [28].

It is known that classification methods are divided into two main groups of techniques, which are cluster analysis and discriminatory analysis. Cluster analysis is one of the techniques without supervision, where information about the classification group is little or no, and the goal is to find groups of data. Places of discriminatory analysis are an organic relationship in the pool of training samples and the main purpose is to build an appropriate classification rule for unknown new samples.

Cluster analysis [29], [30]. Is one of the methods of analysis of multiple variables, and according to this analysis is the division of data into groups (clusters) that have scientific meaning and significance. The cluster analysis is often a step to enter into another analysis that carries goals that depend on the classification of data into clusters that bear the importance of research in the researcher example it summarizes the data. Regardless of whether it is the beginning of a subsequent analysis or classification of data, cluster analysis includes many scientific uses in various

fields and disciplines. The concept of this analysis allows with many of those singular or viewed within a specific cluster that bear close characteristics to the characteristics of observations or data within the regular cluster to it, while those characteristics differ for projects with the characteristics of observations or data in other clusters.

Clustering analysis depends on mathematical methods and algorithms in the classification process that have the ability to determine the extent of homogeneity between the classification vocabularies or the extent of inconsistency between its properties, and then give the researcher an opportunity to explain it within the laws and methods of cluster analysis.

A cluster analysis is a grouping of community units or individuals in order to discover data structure. As individuals within one group are close or seen from one another, but they differ from individuals in other groups, as the results of the cluster analysis provide a known structure that can be used to make hypotheses to interpret the observed data. The cluster analysis which is described as the task of grouping a set of objects in a certain way hence forming a similar group referred to as a cluster is also used in relation to consumption and production of electricity. Literature indicates that cluster analysis has been introduced to help in analysis of electric production and consumption since the invention of smart grid. Advancement of technology has increased the demand of electricity hence forcing producers to scale up production. As such technology which considered more efficient and fast is usually used to ease and perfect the production and consumption. The smart grid was introduced into the energy sphere with the vision of producing and distributing low-carbon electricity more efficiently and reliably. As such, the consumers would be in apposition to manage their use through regulating their usage. In return, costs to both the consumers and the suppliers would be minimized.

Through introduction of the grid, real time communication between the grid operator and the end consumer would be in real time which was intended to help flatten the electricity loads during pick hours. The process also was intended to help increase the options to the end user hence increasing the chances of optimizing the system. According to Rodden and Spence, this idea was referred to as smart load management by the end-user which was proven to be effective. This effectiveness was due to the cluster analysis which according to literature by Anderson its activities are based on those that are reported in time-diaries [31]. Clustering analysis helps or rather allow for assumptions that are empirically grounded in regard to flexibility that is potential [12]. Further literature also indicates that it helps to accurately predict flexibility of customer's behavior. The cluster analysis according to literature is based on actual usage of appliances such as washing machines, cooking equipment's, television sets and many more, it also based on collected activities from the time-diaries [32]. The data collected is usually based on reported activities in real time sequence of how they appear on the time-diaries rather than inferring flexibility indirectly from the load profiles or the time use average values. The

end part is the assignment of data to the electric appliances to help in the characterization of the aggregate activity patterns.

A survey done by Europe Statistics Group indicates that, there is one way in which consumers patterns can be traced which is through the use of data from their time-diaries [23]. This is where the consumer record the activities they are engaged in on daily basis in a certain sequence. The clustering approach therefore involves using the activity sequence deduced from the time-diaries of the consumers [34]. The analysis will then take into consideration the presence of any clusters basing on the similarity of the data gathered. The cluster analysis is used to group the consumers according to their performing activities by using similar timing as well as the duration to assist in the estimation of how the consumers consumes electricity. Some of the activities commonly used include cooking, watching a TV and doing laundry [35]. Research indicates that the consumption of energy is intertwined with people's ways of life and also it is closely associated with how people act at various times of the day and seasons [36]. Cluster analysis in this research helps to determine the time the activities take place and how long they take. As such, the clustering analysis is done for the consumers who are clustered into different socio economic classes. Anderson's research indicates that the lives of consumers are analyzed and described through and explicit reveal of the timing of energy-consuming activities [12]. Anderson is said to introduce new approach of the cluster analysis in this research through raising questions such as whether there are flexible users, and whether or not activities are moveable [12].

Research done by Nicholls and Strangers indicated that, when cluster analysis is done based on activity rather than the consumers, energy consumption in regard to activities causing it vary over the day rendering the approach irrelevant [37]. In the perspective of policies, when the time shift is targeted while analyzing electricity consumption, different groups will be affected differently hence eradicating the possibility of clustering [38]. However the policy means will have to depend on what it is aimed for concerning the flexibility of electricity usage [34]. Clustering in our case is considered effective if individuals adapt to a shift in consumption of electricity from daytime to evenings as compared to vice-versa since the normal patterns are people use more electricity while at home during the evening hours [39]. Clustering is easier while considering consumers since people are easy to predict. People for example prefer to do some things at certain times.

Which can be confirmed by the aggregate activity patterns as indicated in Anderson's research [12]. For example, Anderson conducted a simple survey from which he observed that, during the time, people in New York preferred to do laundry during weekends and at lunch times [12]. However, this data had to be verified to determine its flexibility and whether there is need for complementary methods to verify the validity of the information for future reference [40].

The cluster analysis helps to observe each activity at a time through isolation. The analysis also helps to evaluate whether the activities affects each other. For example, does cooking affect washing or TV time? Does consumption behaviors affect production and if so how? Research by Shove indicated that it is vital to collect the practices that constitute assemblages [41]. Clustering also helps establish the flexibility of the cluster activity by observing the actions of the individual before and after an activity [42].

Individuals or rather consumers are interrelated hence affecting the flexibility and the context of their daily activities. As such, it is important to synchronize and allow time-space coupling between consumers from the same household. To encourage a time-space coupling, data should be collected from each and every consumer even if there are multiple users from one household [43]. Research done by Strengers indicates that the flexibility of a household in regard to energy consumption is related not only to adults but also to everyone in the household including the pets hence determined by the interaction of everyone in the household not excluding animals.

3. Methods

3.1. Fuzzy Time Series Method

The use of fuzzy time series has been used to predict student registration data at the University of Alabama. The concept of fuzzy time series is proposed based on fuzzy set theory, fuzzy logic and approximate reasoning (Song and Chissom, 1993) [18]. Forecasting with fuzzy time series methods can capture patterns from past data to project future data (Song, 1993b) [19], better performance in forecasting real problems.

The various definitions and properties of fuzzy time series forecasting are summarized as follows:

Definition 1: Fuzzy set is an object of classes with a set of membership values. Suppose U is the universe of discourse,

$U = \{u_1, u_2, \dots, u_i\}$ Where u_i is a possible linguistic value of U then a fuzzy set A_i linguistic variable of U is defined by equation 1 following:

$$A_i = \frac{\mu_{A_i}(\mu_1)}{\mu_1} + \frac{\mu_{A_i}(\mu_2)}{\mu_2} + \dots + \frac{\mu_{A_i}(\mu_n)}{\mu_n} \quad (3.1.1)$$

Where μ_{A_i} a membership function is fuzzy set A_i so

$\mu_{A_i} : U \rightarrow [0, 1]$. If the membership is from A_i to u_{A_i} is the degree that is owned by u_j against A_i .

Definition 2: Let $Y(t)$ ($t = \dots, 0, 1, 2, \dots$) subset R^1 , become a universe discourse with the fuzzy set $f_i(t)$ ($i = 1, 2, \dots$) defined and $F(t)$ is a collection of $f_1(t), f_2(t), \dots$, then $F(t)$ is called fuzzy time series defined in $Y(t)$ ($t = \dots, 0, 1, 2, \dots$). From this definition $F(t)$ can be understood as a linguistic variable $f_i(t)$ ($i = 1, 2, \dots$) of the linguistic probability value $F(t)$. Because at different times, the value of $F(t)$ can be

different, $F(t)$ as a fuzzy set is a function.

From time t and universe discourse is different at each time so $Y(t)$ is used for time t (Song and Chissom, 1993).

Definition 3: Suppose $F(t)$ is caused only by $F(t-1)$ and appointed with $F(t-1) \rightarrow F(t)$ then there is Fuzzy Relations between $F(t)$ and $F(t-1)$ expressed by the formula

$$F(t) = F(t-1)R^w(t, t-1) \quad F(t) = F(t-1) \circ R(t, t-1) \quad (3.1.2)$$

Where " \circ " is the Max-Min composition operator. The relation R is called the first order model $F(t)$.

If fuzzy has relation $R(t, t-1)$ of $F(t)$. (t) Is time independent so for different times t_1 and t_2 ,

$R(t_1, t_1-1) = R(t_2, t_2-1)$ So that $F(t)$ called time-invariant fuzzy time series.

Definition 4: If $F(t)$ is produced by several fuzzy sets

$$F(t-n), F(t-n+1) \dots F(t-1)$$

Then the fuzzy relationship is symbolized by:

$$A_i, A_2, \dots, A_{in} \rightarrow A_j \quad \text{Where}$$

$$F(t-n) = A_{i1}, F(t-n+1) = A_{i2} \dots F(t-1) = A_{in} \quad F(t) = A_j$$

And such relationships are called n^{th} order fuzzy time series.

Definition 5: Let $F(t)$ is produced by $F(t-1), F(t-2), \dots$, and $F(t-m)$ ($m > 0$) simultaneously and the relation is time variant then $F(t)$ called fuzzy time series and the relation can be expressed with a formula:

$$F(t) = F(t-1)R^w(t, t-1) \quad (3.1.3)$$

Where $W > 1$ is a time parameter (month or year) that is Affect the forecast $F(t)$.

As per the depiction of the issue, the accompanying determining procedure is proposed:

Step 1: Definition of universal set U containing the interval between the least and biggest variations in whole production and consumption.

Step 2: Division of the universal set U into equal-length intervals containing variation values corresponding to distinctive production and consumption rates to understand how many interval we need.

Step 3: The qualitative description of variation values of total production and consumption as a linguistic variable, that's to say, determining the respective values of linguistic variable or the set of fuzzy sets $F(t)$.

Step 4: Fuzzifying the enter data or the conversion of numerical values into fuzzy values. This operation allows us to mirror the corresponding numerical/qualitative values of qualitative representations of production and consumption rates in the value of membership function.

Step 5: Selection of parameter $W > 1$, corresponding to the time duration prior to the concerned year, calculation of fuzzy Relationships matrix $R^w(T)$ and forecasting of production and consumption in the next month.

Step 6: Defuzzifying the acquired outcomes or conversion of fuzzy values into qualitative values.

3.2. Fuzzy Time Series and Clustering

The proposed clustering algorithm is used to partition universe of discourse into different lengths of intervals [43].

Step 1: Sort the numerical data in ascending sequence, Shown as follows:

$$d_1, d_2, d_3, \dots, d_i, \dots, d_n.$$

Calculate the threshold value for stopping condition of the proposed clustering algorithm, shown as follows:

$$\tau = \frac{\sum_{i=1}^{n-1} (d_{i+1} - d_i)}{n-1}. \quad (3.2.1)$$

Step 2: Put each datum into a cluster, shown as follows

$$\{d_1\}, \{d_2\}, \{d_3\}, \dots, \{d_i\}, \dots, \{d_n\},$$

Where the symbol “{ }” denotes a cluster.

Step 3: Assume that there are p clusters, calculate the cluster center cluster center k of each cluster k as follows:

$$cluster_center_k = \frac{\sum_{j=1}^m d_j}{r},$$

Where dj is the data in Cluster k, r is the number of the data in Cluster k, and $1 \leq k \leq p$.

Calculate the distance $distance_{m,m+1}$ between any two adjacent cluster centers $cluster_center_m$ and $cluster_center_{m+1}$, shown as follows:

$$distance_{m,m+1} = |cluster_center_m - cluster_center_{m+1}|,$$

Where $m = 1, 2, \dots, p-1$ (3.2.2)

Step 4: Find the smallest distance smallest distance:

$$smallest_distance = \min_{m=1, 2, \dots, p-1} (distance_{m,m+1}).$$

Step 5: If smallest distance $< \tau$, then combine the clusters having the smallest distance between them into a cluster and go to Step 3. Otherwise, go to Step 6.

Step 6: Calculate the upper bound $cluster_uBound_m$ of $Cluster_m$ and the lower bound $cluster_lBound_{m+1}$ of $Cluster_{m+1}$:

$$cluster_uBound_m = \frac{cluster_center_m + cluster_center_{m+1}}{2},$$

$$cluster_lBound_{m+1} = cluster_uBound_m, \quad (3.2.3)$$

Where $m = 1, 2, \dots, p-1$. Because there is no previous cluster before the first cluster and there is no next cluster after the last cluster, the lower bound $cluster_lBound_1$ of the first cluster and the upper bound $cluster_uBound_p$ of the last cluster can be calculated as follows:

$$cluster_lBound_1 = 2 \times cluster_center_1 - cluster_uBound_1,$$

$$cluster_uBound_p = 2 \times cluster_center_p - cluster_lBound_p.$$

(3.2.4)

Step 7: Let each cluster $Cluster_k$ form an $interval_k$, which means that the upper bound $cluster_uBound_k$ and the lower bound $cluster_lBound_k$ of the cluster $Cluster_k$ are also the upper bound $interval_uBound_k$ and the lower bound $interval_lBound_k$ of the interval $interval_k$, respectively. Calculate the middle value mid_value_k of the interval $interval_k$ as follows:

$$mid_value_k = \frac{(interval_lBound_k + interval_uBound_k)}{2},$$

Where $1 \leq k \leq p$. (3.2.5)

Step 1: Apply the proposed clustering algorithm to partition the universe of discourse.

Step 2: Assume that there are n intervals u_1, u_2, \dots, u_n obtained in Step 1, and then define linguistic terms A_1, A_2, \dots, A_n represented by fuzzy sets, shown as follows:

$$A_1 = 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_n,$$

$$A_2 = 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{n-1} + 0/u_n,$$

$$A_3 = 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{n-1} + 0/u_n,$$

.

.

$$A_n = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{n-1} + 1/u_n.$$

Step 3: Fuzzily each historical datum into a fuzzy set. If the datum is belonging to u_i , then the datum is fuzzified into A_i , where $1 \leq i \leq n$.

Step 4: Construct the fuzzy logical relationship based on the fuzzified data obtained in Step 3.

(Note: If the first order fuzzy time series is used and the fuzzified values of time t-1 and t are A_j and A_K , respectively, then construct the fuzzy logical relationship “ $A_j \rightarrow A_K$ ”, where “ A_j ” and “ A_K ” are called the current state and the next state of the fuzzy logical relationship. If the nth order fuzzy time series is used and the fuzzified values of time t-n... t-2, t-1 and t are $A_{j,n}, \dots, A_{j,2}, A_{j,1}$ and A_K , respectively, then construct the Fuzzy logical relationship “ $A_{j,n}, \dots, A_{j,2}, A_{j,1} \rightarrow A_K$ ”, where “ $A_{j,n}, \dots, A_{j,2}, A_{j,1}$ ” and “ A_K ” are called the current state and the next state of the nth order fuzzy logical relationship).

Based on the current state of the fuzzy logical relationships, let the fuzzy logical relationships having the same current state to form a fuzzy logical relationship group.

Step 5: Calculate the forecasted output at time t by using the following principles:

Principle 1: If the fuzzified values at time t-n... t-2, and t-1 are $A_{j,n}, \dots, A_{j,2}$, and $A_{j,1}$, respectively, and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, shown as follows:

$$A_{j,n}, \dots, A_{j,2}, A_{j,1} \rightarrow A_k,$$

Then the forecasted value of time t is m_k , where m_k are the middle value of the interval u_k and the maximum membership value of A_k occurs at interval u_k .

Principle 2: If the fuzzified values at time $t-n \dots t-2$, and $t-1$ are $A_{j,n}, \dots, A_{j,2}$, and $A_{j,1}$, respectively, and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, shown as follows:

$$A_{j,n}, \dots, A_{j,2}, A_{j,1} \rightarrow A_{k1}(x_1), A_{k2}(x_2), \dots, A_{kp}(x_p),$$

Then the forecasted value of time t is calculated as follows:

$$\frac{x_1 \times m_{k1} + x_2 \times m_{k2} + \dots + x_p \times m_{kp}}{x_1 + x_2 + \dots + x_p}, \quad (3.2.6)$$

where x_i denotes the number of fuzzy logical relationships " $A_{j,n}, \dots, A_{j,2}, A_{j,1} \rightarrow A_{ki}$ " in the fuzzy logical relationship group, $1 \leq i \leq p$; m_{k1}, m_{k2}, \dots , and m_{kp} are the middle value of the intervals u_{k1}, u_{k2}, \dots , and u_{kp} , respectively, and the maximum membership values of A_{k1}, A_{k2}, \dots , and A_{kp} occur at interval u_{k1}, u_{k2}, \dots , and u_{kp} , respectively.

Principle 3: If the fuzzified values at time $t-n \dots t-2$, and $t-1$ are $A_{j,n}, \dots, A_{j,2}$, and $A_{j,1}$, respectively, and there is only one fuzzy logical relationship in the fuzzy logical relationship groups, shown as follows:

$$A_{j,n}, \dots, A_{j,2}, A_{j,1} \rightarrow \#,$$

Then the forecasted value of time t is calculated as follows:

$$\frac{m_{j,n} + \dots + m_{j,2} + m_{j,1}}{n}, \quad (3.2.7)$$

where $m_{j,n}, \dots, m_{j,2}$ and $m_{j,1}$ are the middle values of the intervals $u_{j,n}, \dots, u_{j,2}$ and $u_{j,1}$, respectively, and the maximum membership values of $A_{j,n}, \dots, A_{j,2}$ and $A_{j,1}$ occur at intervals $u_{j,n}, \dots, u_{j,2}$ and $u_{j,1}$, respectively.

3.3. Proposed Method for Predicting Using Fuzzy Time Series and Clustering

Through our studies Fuzzy time series and clustering Time series analysis is an important statistical topic that deals with the behavior of phenomena and their interpretation over specific periods. The objectives of time series analysis can be summed up by obtaining an accurate description of the special features of the process from which the time series is generated, building a model to explain the behavior of the time series and using the results to predict the behavior of the series in the future, in addition to controlling the process

from which the time series is generated by examining what can happen when some change Form parameters. To achieve this, a thorough analytical study of time-series models based on statistical and mathematical methods are required.

Our proposal includes the same previous steps Fuzzy Time Series and Clustering but we have specials steps and they are as follows:

First Step: the data will take 60 months from January 2015 to December 2019 not the same in the section 3.2 we take the period from December 2015 to December 2019 because our goal in this paper prediction next months and months.

If you use the same steps in the section 3.2 we cannot prediction the future just we predict one value.

Second step: the method which used in section 3.2 used.

Definition three in fuzzy time series in section 3.1 in our proposal method we will use we use time-invariant fuzzy time series in section 3.1.

Third step: the method which used in section 3.2 take If A_j is not related to any other group, i.e. $A_j \rightarrow \theta$ where θ is the empty group, and the highest degree of affiliation with A_j is in the period u_K , then the results of the prediction are equal to the middle of the period u_K .

But we us this relation Mean of monthly variation

This calculated for every month by the equation below

$$\frac{\text{Actual value} - \text{previous value}}{\text{previous value}} \quad (3.3.1)$$

After that calculate the average mean of monthly variation for all months then it A_j is not related to any other group.

We will get Forecasted Value of the fuzzy membership by mid value $\times (1 + \text{average mean of monthly variation})$ (3.2.2).

Fourth step: the difficult step and final step how many order equation I will use first, second, third order equation.

In the method in section 3.2 use first order equation but in our Proposed Method twelve order equation but the condition here you must choose the data equal for example we choose data from January 2015 to December 2019 has 60 elements And we need prediction for 12 months so the data will be equal $60/12=5$.

4. Model Evaluation

The process of evaluating models is intended to evaluate the field suitability of the model for the pattern in which the series data is running or the accuracy of the model in predicting the values of the current and future series, and there are many measures of the suitability of the model all depend on the degree of error, which is the difference between the actual value of the series at a specific time And the string value that the model expected at that time. In this study, we will rely on the following methods to compare the two models used in this paper to find out which one is more accurate in prediction.

4.1. Mean Absolute Percentage Error (MAPE)

$$MAPE = 100 \sum_{i=1}^n \left[\frac{|Y_i - F_i|}{Y_i} \right] / n \quad (4.1.1)$$

The scale that eliminates the problem that negative error values cancel out positive values, and does not amplify the error by squaring, as happens in the sum of squares of errors. It is possible to compare models across different series. It is called "Average absolute values of error ratios".

4.2. Mean Absolute Deviation (MAD)

$$MAD = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (4.2.1)$$

The mean absolute deviation of a dataset is the average distance between each data point and the mean. It gives us an idea about the variability.

5. Data Collecting and Processing

The data used in this research was gathered by the <https://www.ceicdata.com/en> from 2015-01 to 2019-12. It contained the data of production and consumption electricity in china. in a dataset.

6. Experimental Results and Discussion

6.1. Result Fuzzy Time Series Method and Discussion

The first step: The dynamics of total production over May 2015-December 2020 years (input data for retrospective forecast) and variation in total production between every next and previous month. Variation for the current month is understood to be the difference between the sizes of production in current and previous months. For example, variation for Aug 2015 is equal = Aug 2015 – July 2015 then 3738.04336 - 3220.84444 = 517.19892. To define a universal set U, first of all, the smallest and greatest variation values must be found over the period [Dec 2015, Dec2019], later, to ensure the smoothness of boundaries of the interval, adequate values D1, D2 (positive figures are selected. After that, the universal set U can be defined as $U : U = [V_{min} - D_1, V_{max} + D_2]$, where $V_{min} = -6567.178$ is the smallest variation (Jan 2019), $V_{max} = 1130.145$ is the greatest variation (Mar 2019), $D_1 = 332.8$, $D_2 = 369.85$. Thus, the universal set U will be as follows: $U = \{-6900, 1500\}$.

The second step: The universal set U must be divided into several equal intervals. In our case, this set U is divided into seven equally- length intervals: $u_1 = [-6900, -5700]$, $u_2 = [-5700, -4500]$, $u_3 = [-4500, -3300]$, $u_4 = [-3300, -2100]$, $u_5 = [-2100, -900]$, $u_6 = [-900, 300]$, $u_7 = [300, 1500]$. Basically we calculated the length of the interval U which is $(1500) - (-6900) = 8400$ and divided it by $7: 8400/7=1200$, then we built our small intervals: (example: $u_1 = [-6900, -5700]$) which has 1200 as magnitude.

If we take into account the fact that forecasting with fuzzy time series exhibits the least average error, it's necessary to find the middle points of the intervals: $um_1 = -6300$, $um_2 = -5100$, $um_3 = -3900$, $um_4 = -2700$, $um_5 = -1500$, $um_6 = -300$, $um_7 = 900$.

The third step: Fuzzy sets are defined on the universal set U. In this case "the variation in total production" is a linguistic variable that assumes the following linguistic values: A1=(very low level production electric (VLLPE)); A2=(low level production electric (LLPE)); A3=(changeless production electric (CPE)); A4=(moderate production electric (MPE)); A5=(normal-level production electric (NLPE)); A6=(high-level production electric (HLPE)); A7=(very high-level production electric (VHLPE)). To every linguistic value here corresponds a fuzzy variable which, according to a certain rule is assigned against a corresponding fuzzy set determining the meaning of this variable.

For example, the linguistic value "very-low-level production electric" is given by the fuzzy variable $\langle VLLPE, [-6900, -5700], A1 \rangle$, where A1 is a fuzzy set defined on the domain $[-6900, -5700]$ of the universal set U. See example (3).

The fuzzy set A1, A2... A7 is defined on the universal set U by the following formula (6.1.1):

$$\mu_{A_i}(u_i) = \frac{1}{1 + \left[C \cdot (U - u_m^i) \right]^2} \quad (6.1.1)$$

u_m^i is the middle point of the corresponding interval in (1); C is a constant. C is chosen in such a way that it ensures the conversion of definite quantitative values into fuzzy values or their belonging to the interval. (In our case $C=0.0001$); $A_i = (\mu_{A_i}(u_i) / u_i, u_i \in [0, 1])$ is a fuzzy set

If the value of the variable U in formula (6.1.1) is accepted as the middle point of the corresponding interval, the fuzzy set A_i ($i=1 \dots 7$) will be defined as follows:

$A1 = \{(1/u_1), (0.61/u_2), (0.27/u_3), (0.15/u_4), (0.10/u_5), (0.06/u_6), (0.04/u_7)\}$

$A2 = \{(0.61/u_1), (1/u_2), (0.61/u_3), (0.27/u_4), (0.15/u_5), (0.10/u_6), (0.06/u_7)\}$

$A3 = \{(0.27/u_1), (0.61/u_2), (1/u_3), (0.61/u_4), (0.27/u_5), (0.15/u_6), (0.10/u_7)\}$

$A4 = \{(0.15/u_1), (0.27/u_2), (0.61/u_3), (1/u_4), (0.61/u_5), (0.27/u_6), (0.15/u_7)\}$

$A5 = \{(0.10/u_1), (0.15/u_2), (0.27/u_3), (0.61/u_4), (1/u_5), (0.61/u_6), (0.27/u_7)\}$

$A6 = \{(0.06/u_1), (0.10/u_2), (0.15/u_3), (0.27/u_4), (0.61/u_5), (1/u_6), (0.61/u_7)\}$

$A7 = \{(0.04/u_1), (0.06/u_2), (0.10/u_3), (0.15/u_4), (0.27/u_5), (0.61/u_6), (1/u_7)\}$

The fourth step: This step consists of the fuzzification of the variation calculated at the first step. This time, if $B_u, B_u \in y_j$ is a variation for the i-th month, then membership function for $\mu(y_b)$ is calculated by means of formula (6.1.1) by holding valid the equality $Y = B_u$, that's to

say, by separating the interval, to which belongs the considered variation, from the universal set U.

Here, $A^{mn} - t = mn$ is a fuzzy set of the corresponding variation for the year $t = m \times n$ where May2015 < t Dec2019.

The fifth step: We must select a basis w ($1 < w < l$, where l is the number of months, prior to the current month included in experimental evaluation). Resting on the basis W or the past months, we calculate a fuzzy relationship matrix $R^w(t)$ by means of which is given a forecast. For this purpose, after the selection of w , we establish an operation matrix $i \times j$ $O^w(t)$ (here i is the number of rows, Criteria matrix $1 \times j$ $K(t)$ (a row matrix corresponding to fuzzy variation in total

population for the month $t-1$). For example, by assuming that $w=7$, we can define the operation matrix 6×7 $O^7(T)$ (which is the matrix of fuzzy variations in total production electric over the months $t-2, t-3, t-4, t-5, t-6, t-7$) and the criteria matrix 1×7 $K(t)$ (which is the fuzzy variation matrix for the month $t-1$). Thus for $w=7$, the previous 8 months data are utilized (the total production electric of the $(t-8)$ month must be known to find variation of the $(t-7)$ month).

At last, for example, in order to forecast the total production electric for July 2017, the operation matrix $O^7(t)$ will be established as follows

$O^7 =$ (July 2017) for

fuzzy variation in total population for the Dec – 2016
fuzzy variation in total population for the Jan – 2017
fuzzy variation in total population for the Feb – 2017
fuzzy variation in total population for the Mar – 2017
fuzzy variation in total population for the Apr – 2017
fuzzy variation in total population for the May – 2017

$O^7(1990) =$

	VLLPE	LLPE	CPE	MPE	NLPE	HLPE	VHLPE
Month	Fuzzy variations						
Dec-16	0.00017	0.000169	0.000169964	0.000170497	0.000171	0.000172	0.000172
Jan-17	3.12031E-06	3.11899E-06	3.11767E-06	3.11635E-06	3.11503E-06	3.11371E-06	3.11239E-06
Feb-17	0.72	0.79	0.87	0.93	0.98	1.00	0.99
Mar-17	0.000100139	0.00010038	0.000100622	0.000100865	0.000101108	0.000101352	0.000101598
Apr-17	0.000423512	0.000425611	0.000427725	0.000429856	0.000432002	0.000434165	0.000436344
May-17	0.000392229	0.000394099	0.000395983	0.000397881	0.000399792	0.000401717	0.000403656

$K(\text{Jul-2017}) = [\text{fuzzy variation in total production electric for the Jun2017-th month}] - [\text{fuzzy variation in total population for Jun-2017}]$, That is to say

	VLLPE	LLPE	CPE	MPE	NLPE	HLPE	VHLPE
Months	Fuzzy variations						
Jun-17	0.000356733	0.000358355	0.000359988	0.000361633	0.000363289	0.000364956	0.000366635

According to the method, the relationship matrix $R(t)$ is calculated at the next step

$$R(t)[i, j] = O^w(t)[i, j] \cap K(t)[j],$$

$$R(t) = O^w(t) \times K(t) = \begin{pmatrix} R_{11} & R_{12} & \dots & R_{1j} \\ \vdots & \ddots & & \vdots \\ R_{i1} & R_{i2} & \dots & R_{ij} \end{pmatrix} \quad (6.1.2)$$

Where $O^w(T)$ is an operation matrix; $P(t)$ is matrix of fuzzy sets, \otimes -is an operation $\min(\cap)$. Later there is defined the forecasted value $F(t)$ for the t year in a fuzzy form as follows.

$$F(t) = \left[\text{Max}(R_{11}, R_{21}, \dots, R_{i1}) \text{Max}(R_{1j}, R_{2j}, \dots, R_{ij}) \right]$$

In our case $1 \leq i \leq 6, 1 \leq j \leq 7$ (6.1.3)

R(Jul-17)	0.000168906	0.000169434	0.000169964	0.000170497	0.000171033	0.000171571	0.000172112
	3.12031E-06	3.11899E-06	3.11767E-06	3.11635E-06	3.11503E-06	3.11371E-06	3.11239E-06
	0.72	0.79	0.87	0.93	0.98	1.00	0.99
	0.000100139	0.00010038	0.000100622	0.000100865	0.000101108	0.000101352	0.000101598
	0.000423512	0.000425611	0.000427725	0.000429856	0.000432002	0.000434165	0.000436344
	0.000392229	0.000394099	0.000395983	0.000397881	0.000399792	0.000401717	0.000403656

Finally, the results obtained from population forecast for the July 2017 will be as follows.

F(Jul-17)	0.72	0.79	0.87	0.93	0.98	1.00	0.99
-----------	------	------	------	------	------	------	------

Forecasting results for other years are calculated in an same manner.

The sixth step: to fuzzify the obtained results of the 5-th step, the following formula is proposed

$$V(t) = \frac{\sum_{i=1}^7 \mu_t(u_i) \cdot u_m^i}{\sum_{i=1}^7 \mu_t(u_i)} \quad (6.1.4)$$

Where $\mu_T(y_\mu)$ is the calculated value of membership function for the forecast year t, u_m^i are the middle points of intervals. For example, after calculating F (Jul-17) = -2.511164519, that is to say, anticipated production electric July 2017 equals to -2.511164519. In orders to estimate the forecasted total production electric for July 2017, we must add the calculated production electric to the total production for the June 2016. In other words

$$N(\text{July 2017}) = 2959.83 + (-2.511164519) = 2957.32.$$

Table 6.1.1 and Table 6.1.2 below show the actual and forecasted values of production and consumption electric for period from January 2016 to December 2019 in GWh the result has been rounded to the nearest integer.

Table 6.1.1. Actual and forecasted values of production electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Fuzzy Time Series Method	2016	435	5812	435	433	1355	435	1799	1353	2268	1796	2759	2265	3312	2757	3877	3310	4373	3875	4865	4371	5370	4862	6133	5368
	2107	466	6131	466	463	1459	466	1938	1456	2437	1936	2960	2434	3570	2957	4166	3567	4689	4163	5194	4687	5712	5192	6604	5709
	2018	523	6602	523	520	1576	523	2088	1574	2636	2085	3195	2634	3837	3192	4480	3835	5036	4478	5582	5034	6163	5579	7112	6160
	2019	545	7109	545	542	1675	545	2220	1672	2781	2217	3367	2778	4030	3365	4703	4028	5297	4700	5874	5294	6480	5872	7503	6477

Table 6.1.2. Actual and forecasted values of consumption electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Fuzzy Time Series Method	2016	486	5799	845	493	1290	876	1732	1350	2189	1807	2662	2280	3167	2773	3678	3326	4134	3889	4584	4386	5049	4875	5802	5382
	2107	487	6127	876	484	1352	936	1809	1444	2282	1928	2776	2424	3329	2948	3892	3555	4389	4155	4878	4686	5385	5199	6130	5731
	2018	487	6361	936	597	1446	1055	1931	1585	2426	2107	2951	2660	3558	3227	4157	3875	4689	4527	5202	5104	5733	5653	6364	6217
	2019	600	6898	1055	615	1588	1106	2109	1677	2663	2230	3229	2797	3878	3396	4530	4063	5106	4740	5655	5342	6220	5921	6900	6512

6.2. Result Fuzzy Time Series and Clustering Method and Discussion

The Proposed Method using the First Order Fuzzy Time Series.

[Step 1] Apply the proposed clustering algorithm to partition UoD into different lengths of intervals.

[Step 2] Sorting the numerical data: 5814.57, 435.09, 435.09, 1355.14, 1798.59, 2267.61, 2759.49, 3312.05, 3877.16, 4373.23, 4864.69, 5370.08, 6133.16, 465.77, 465.77, and 1458.72.....etc.

Calculate the threshold τ for stopping condition of the proposed clustering algorithm:

$$\tau = \frac{\sum_{i=1}^{49} (d_{i+1} - d_i)}{49} = 144.251908.$$

[Step 3] Put each datum in a cluster, shown as follows: {5814.57}, {435.09}, {435.09}, {1355.14}, {1798.59},

{2267.61}, {2759.49}, {3312.05}, {3877.16}, {4373.23}, {4864.69}, {5370.08}, {6133.16}, {465.77}, {465.77}, {1458.72}, {1938.24}, {2436.77}, {2959.83}, {3569.76}, {4165.94}, {4689.14},.....etc..

[Step 4] Based on Eq. (2), calculate each cluster center $cluster_center_k$ $1 \leq k \leq 49$, Based on Eq. (3.2.2), calculate the $distance_{m,m+1}$, $1 \leq m \leq 49$, shown as follows:

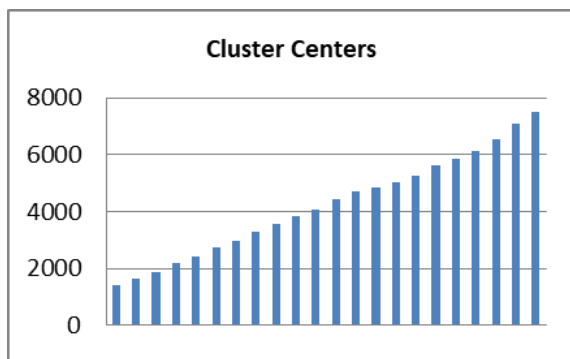
Find the smallest distance $smallest_distance$, i.e., 30.68 (the distance $distance_2$, 3 between $cluster_center_2$ and $cluster_center_3$).

Table 6.2.1 show the Distance between clusters for production electric

[Step 5] Because the $smallest_distance < \tau$, i.e., $30.68 < 144.2519077$ is true, then $center_2$ (i.e., {465.77}) and $center_3$ (i.e., {522.73}) are combined into one cluster (i.e., {465.77, 522.73}), and go to Step 2 the iterations of Step 3 to Step 5 are repeatedly done until the condition “ $smallest_distance < \tau$ ” is false.

Table 6.2.1. Distance between clusters

$cluster_center_1$	492.0428813	
$cluster_center_2$	1406.92839	$distance_{1,2} = 914.8855088$
$cluster_center_3$	1625.505	$distance_{2,3} = 218.57661$
$cluster_center_4$	1868.415	$distance_{3,4} = 242.91$
$cluster_center_5$	2191.686667	$distance_{4,5} = 323.2716667$
$cluster_center_6$	2436.77	$distance_{5,6} = 245.0833333$
$cluster_center_7$	2725.51	$distance_{6,7} = 288.74$
$cluster_center_8$	2959.83	$distance_{7,8} = 234.32$
$cluster_center_9$	3291.286667	$distance_{8,9} = 331.4566667$
$cluster_center_{10}$	3569.76	$distance_{9,10} = 278.4733333$
$cluster_center_{11}$	3857.245	$distance_{10,11} = 287.485$
$cluster_center_{12}$	4097.995	$distance_{11,12} = 240.75$
$cluster_center_{13}$	4426.65	$distance_{12,13} = 328.655$
$cluster_center_{14}$	4695.87	$distance_{13,14} = 269.22$
$cluster_center_{15}$	4864.69	$distance_{14,15} = 168.82$
$cluster_center_{16}$	5036.17	$distance_{15,16} = 171.48$
$cluster_center_{17}$	5287.08	$distance_{16,17} = 250.91$
$cluster_center_{18}$	5646.725	$distance_{17,18} = 359.645$
$cluster_center_{19}$	5844.371631	$distance_{18,19} = 197.6466309$
$cluster_center_{20}$	6147.880026	$distance_{19,20} = 303.5083952$
$cluster_center_{21}$	6542.0435	$distance_{20,21} = 394.1634739$
$cluster_center_{22}$	7111.773	$distance_{21,22} = 569.7295$
$cluster_center_{23}$	7503.43	$distance_{22,23} = 391.657$

**Figure 1.** Cluster centers of production electric

The final clustering results are in column 2 of the table 6.2.1 show how get it.

[Step 6] Based on Equation (3.2.3), the upper bound and lower bound of each, $1 \leq k \leq 11$. For example:

$$\begin{aligned}
 cluster_uBound_{13} &= \frac{cluster_center_{12} + cluster_center_{13}}{2} \\
 &= \frac{3207.68722 + 3339.665}{2} \\
 &= 3454.7125 \\
 cluster_lBound_{13} &= cluster_uBound_{12} \\
 &= 3273.67611
 \end{aligned}$$

Because there is no previous cluster before $Cluster_1$, the lower bound of $cluster_lBound_1$ of $Cluster_1$ is calculated using Eq. (3.2.4) and because there is no next cluster after the last cluster, i.e., $Cluster_{30}$ the upper bound

$cluster_uBound_{30}$ is calculated using.

$$\begin{aligned} cluster_lBound_1 &= 2 \times cluster_center_1 - cluster_uBound_1 \\ &= 2 \times 491.774505 - 673.930253 \\ &= 309.618758. \end{aligned}$$

$$\begin{aligned} cluster_uBound_{30} &= 2 \times cluster_center_{30} - cluster_lBound_{30} \\ &= 2 \times 7503.43 - 7307.6015 \\ &= 7699.2585 \end{aligned}$$

[Step 7] Let each $Cluster_k$ form an $interval_k$ and calculate the middle value using Eq. (3.2.5). The table 6.2.2 below shows Interval Generations from the Clusters and Final intervals from clusters show in table 6.2.3.

Table 6.2.2. The Interval Generations from the Clusters

Cluster center	Lower Band	Upper Band	Mid value
492.0428813	34.60012688	949.4856356	492.0428813
1406.92839	949.4856356	1516.216695	1232.851165
1625.505	1516.216695	1746.96	1631.588348
1868.415	1746.96	2030.050833	1888.505417
2191.686667	2030.050833	2314.228333	2172.139583
2436.77	2314.228333	2581.14	2447.684167
2725.51	2581.14	2842.67	2711.905
2959.83	2842.67	3125.558333	2984.114167
3291.286667	3125.558333	3430.523333	3278.040833
3569.76	3430.523333	3713.5025	3572.012917
3857.245	3713.5025	3977.62	3845.56125
4097.995	3977.62	4262.3225	4119.97125
4426.65	4262.3225	4561.26	4411.79125
4695.87	4561.26	4780.28	4670.77
4864.69	4780.28	4950.43	4865.355
5036.17	4950.43	5161.625	5056.0275
5287.08	5161.625	5466.9025	5314.26375
5646.725	5466.9025	5745.548315	5606.225408
5844.371631	5745.548315	5996.125828	5870.837072
6147.880026	5996.125828	6344.961763	6170.543796
6542.0435	6344.961763	6826.90825	6585.935007
7111.773	6826.90825	7307.6015	7067.254875
7503.43	7307.6015	7699.2585	7503.43

Table 6.2.3. Final intervals from clusters

No	Intervals
U ₁	[34.60012688, 949.4856356)
U ₂	[949.4856356, 1516.216695)
U ₃	[1516.216695, 1746.96)
U ₄	[1746.96, 2030.050833)
U ₅	[2030.050833, 2314.228333)
U ₆	[2314.228333, 2581.14)
U ₇	[2581.14, 2842.67)
U ₈	[2842.67, 3125.558333)

U ₉	[3125.558333, 3430.523333)
U ₁₀	[3430.523333, 3713.5025)
U ₁₁	[3713.5025, 3977.62)
U ₁₂	[3977.62, 4262.3225)
U ₁₃	[4262.3225, 4561.26)
U ₁₄	[4561.26, 4780.28)
U ₁₅	[4780.28, 4950.43)
u ₁₆	[4950.43, 5161.625)
u ₁₇	[5161.625, 5466.9025)
U ₁₈	[5466.9025, 5745.548315)
U ₁₉	[5745.548315, 5996.125828)
U ₂₀	[5996.125828, 6344.961763)
U ₂₁	[6344.961763, 6826.90825)
U ₂₂	[6826.90825, 7307.6015)
U ₂₃	[7307.6015, 7699.2585)

Table 6.2.4. FLR of the first order Fuzzy Time Series

Group 1: $A_1 \rightarrow A_1(4), A_1 \rightarrow A_2(2), A_1 \rightarrow A_3(2),$
Group 2: $A_2 \rightarrow A_4, A_4$
Group 3: $A_3 \rightarrow A_5, A_5$, Group 4: $A_4 \rightarrow A_5, A_6$
Group 5: $A_5 \rightarrow A_7(3)$, Group 6: $A_6 \rightarrow A_8$
Group 7: $A_7 \rightarrow A_9(3)$, Group 8: $A_8 \rightarrow A_{10}$
Group 9: $A_9 \rightarrow A_{11}(2), A_{12}$, Group 10: $A_{10} \rightarrow A_{12}$
Group 11: $A_{11} \rightarrow A_{13}(2)$, Group 12: $A_{12} \rightarrow A_{14}, A_{14}$
Group 13: $A_{13} \rightarrow A_{15}, A_{16}$, Group 14: $A_{14} \rightarrow A_{17}, A_{17}$
Group 15: $A_{15} \rightarrow A_{17}$, Group 16: $A_{16} \rightarrow A_{18}$
Group 17: $A_{17} \rightarrow A_{18}, A_{19}, A_{20}$
Group 18: $A_{18} \rightarrow A_{20}, A_{21}$
Group 19: $A_{19} \rightarrow A_1, A_{21}$, Group 20: $A_{20} \rightarrow A_1, A_{22}$
Group 21: $A_{21} \rightarrow A_1, A_{23}$, Group 22: $A_{22} \rightarrow A_1$
Group 23: $A_{23} \rightarrow \theta$

[Step 1] Define the linguistic term A_1, A_2 , and A_{30} , shown as follows:

$$\begin{aligned} A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + \dots + 0/u_{10} + 0/u_{11}, \\ A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + \dots + 0/u_{10} + 0/u_{11}, \\ A_3 &= 0/u_1 + 0.5/u_2 + 1/u_3 + 0.5/u_4 + \dots + 0/u_{10} + 0/u_{11}, \end{aligned}$$

.

.

$$A_{30} = 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + \dots + 0.5/u_{10} + 1/u_{11}$$

[Step 2] Fuzzify each datum that is belonging to u_i , where

$1 \leq i \leq 11$ into A_i .

[Step 3] Obtain the fuzzy logical relationships (FLR) of the first order fuzzy time series Table (6.2.4). Let the FLR having the same current state to form a FLR group (FLRG).

[Step 4] below in Table 6.2.5 we calculate the forecasted value of the fuzzy memberships by using the mean of mid values weighted by their corresponding repetitions. Again we will give ex: The state A_2 comes after state A_1 2 times. So the forecasted value of A_2 is the mid value of A_2 two times divided by 2 (1888.505417×2)/2. Not clear enough? We will give another example.

The state A_8 comes after state A_7 , A_8 once and A_{10} once. So the forecasted value of A_8 is the Midvale of A_8 and the Midvale of A_4 divided by the total number of repetitions ($3572.012917 \times 1 + 4119.97125 \times 1$)/2.

[Step 5] Calculate the forecasting value. In column I we start forecasting our data by using states from the previous year. How we do this? Let's see some examples:

We forecast the value of February 2016 by looking at the state of January 2016 and we affect the corresponding forecasted value which is (962).

We forecast the value of July 2016 by looking at the state of Jun 2016 and we affect the corresponding forecasted value from column C which is (3178), Table 6.2.6 and Table 6.2.7 below show the actual and forecasted values of production and consumption electric for period from January 2016 to December 2019 in GWh. the result has been rounded to the nearest integer.

Table 6.2.5. Fuzzified production electric

	Forecasted Value of the fuzzy membership
A1	962.1313188
A2	1888.505417
A3	2172.139583
A4	2309.911875
A5	2711.905
A6	2984.114167
A7	3278.040833
A8	3572.012917
A9	3937.03125
A10	4119.97125
A11	4411.79125
A12	4670.77
A13	4960.69125
A14	5314.26375
A15	5314.26375
A16	5606.225408
A17	5882.535425
A18	6378.239401
A19	3779.648878
A20	3779.648878
A21	3997.736441
A22	492.0428813
A23	962.1313188

Table 6.2.6. Actual and forecasted values of production electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Fuzzy Time Series and Clustering Method	2016	435	3780	435	962	1355	962	1799	1889	2268	2310	2759	2712	3312	3278	3877	3937	4373	4412	4865	4961	5370	5314	6133	5883
	2107	466	3780	466	962	1459	962	1938	1889	2437	2310	2960	2984	3570	3572	4166	4120	4689	4671	5194	5314	5712	5883	6604	6378
	2018	523	3998	523	962	1576	962	2088	2172	2636	2712	3195	3278	3837	3937	4480	4412	5036	4961	5582	5606	6163	6378	7112	3780
	2019	545	492	545	962	1675	962	2220	2172	2781	2712	3367	3278	4030	3937	4703	4671	5297	5314	5874	5883	6480	3780	7503	3998

Table 6.2.7. Actual and forecasted values of consumption electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Fuzzy Time Series and Clustering Method	2016	495	3561	876	1028	1352	1506	1809	1896	2282	2315	2776	2717	3329	3287	3892	3948	4389	4448	4878	4906	5385	5394	6130	6021
	2107	487	3717	936	1028	1446	1506	1931	1896	2426	2315	2951	2992	3558	3580	4157	4124	4689	4541	5202	5266	5733	5654	6364	6249
	2018	600	550	1055	1028	1588	1506	2109	2178	2663	2717	3229	3287	3878	3948	4530	4448	5106	5137	5655	5654	6220	6249	6900	3717
	2019	617	550	1106	1028	1680	1506	2233	2178	2799	2717	3398	3287	4065	3948	4742	4541	5344	5266	5923	6021	6514	3561	7226	7226

6.3. Result Proposed Method for Predicting Using Fuzzy Time Series and Clustering Discussion

The Proposed Method using Fuzzy Time Series and Clustering the same steps but we will use data from January 2015 to December 2019 in our proposal method we will use time-invariant fuzzy time series in section 3.1, the first step and second step The same in section 4.2 and will start from step three in proposed method.

[Step 3] Obtain the fuzzy logical relationships (FLR) of the first order fuzzy time series. Let the FLR having the same current state to form a FLR group (FLRG).

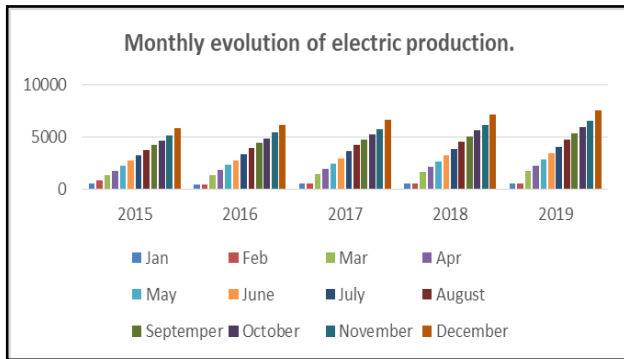


Figure 2. Monthly evolution of electric production

In Figure 2, this histogram gives us a clear idea about the distribution of our data, we can see that

In every year electrical production rises continually through the months of the year to reach its peak in December and fall abruptly in the next month of the year. Because of its shape that looks the same every year we say that our data is affected by seasonality with a yearly span. You can figure out that there are other kind of data which has a span of 6 months (semester) or 3 months (a quarter)... So we would like that our model reflect this property.

Here we calculated the conditional mean using the mean of monthly variations from January to December the table 6.3.1 below indicate how get it, this way we can tell that in average the data should evolve with 6% in the next year. We will not forecast all data this way but this should be a good

estimation for our problematic states as $A_j \rightarrow \theta$.

We can see in Figure 2 that the production of electricity is increasing throughout the years. So we would like that our forecasting to be affected too. It's visually clearer below in Figure 3 (we can also check the mean of variations in row 8 which shows that difference in trend)

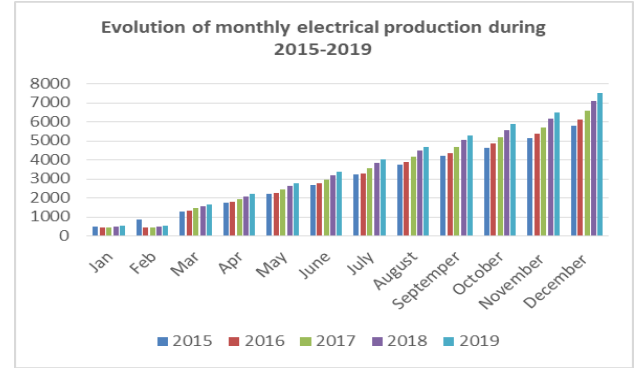


Figure 3. Monthly evolution of electric production

[Step 4] we calculate the forecasted value of the fuzzy memberships by using the mean of mid values weighted by their corresponding repetitions from table 6.3.2. Again we will give ex:

The state A1 comes after state A1 7 times. So the forecasted value of A1 is the mid value of A1 seven times divided by 7 ($491, 774505 \times 7 / 7$). Not clear enough? We will give another example.

The state A3 comes after state A3 once and A4 once. So the forecasted value of A3 is the Midvale of A3 and the Midvale of A4 divided by the total number of repetitions ($(1259, 747445 \times 1 + 1527, 9773475 \times 1) / 2$).

The state that doesn't have a state that comes after 12 months are basically the states that were realized in their first time in year 2019, so we'll just suppose how they should evolve in the next year (6%).

The state $A_{27} \rightarrow \theta$ which state come after state A26

the forecasted value of A27 by equation (3.2.2)

Mid value (1+ average mean of monthly variation)

$6427.901757 \times (1 + 0.06) = 6813.575862$

Table 6.3.1. The mean of monthly variation

Month	2016	2017	2018	2019	Mean of monthly variation
Jan	-0.1133367	0.07051120	0.12229343	0.041838443	0.101397554
Feb	-0.4917724	0.07051120	0.12229343	0.041838443	-0.064282333
Mar	0.03424081	0.07643746	0.08058434	0.062470262	0.063433222
Apr	0.02232403	0.07764415	0.07708539	0.063300841	0.060088605
May	0.02206442	0.07459836	0.08180911	0.054929214	0.058350281
June	0.01861736	0.07260037	0.07929509	0.054076813	0.056147411
July	0.02831728	0.07780981	0.07495461	0.05022242	0.057826034
August	0.03721643	0.07448235	0.07540435	0.049671099	0.059193559
September	0.04150504	0.07223722	0.07400717	0.051737729	0.059871793
October	0.04591796	0.06778232	0.07454138	0.052411213	0.060163221
November	0.04768296	0.06363778	0.07892055	0.051445818	0.060421781
December	0.05479108	0.07684243	0.07681581	0.05507164	0.065880245

Table 6.3.2. FLR of the first order Fuzzy Time Series

Group 1: $A_1 \rightarrow A_1 (7)$	Group 2: $A_2 \rightarrow A_1$
Group 3: $A_3 \rightarrow A_3, A_4$	Group 4: $A_4 \rightarrow A_4, A_5$
Group 5: $A_5 \rightarrow A_5, A_6$	Group 6: $A_6 \rightarrow A_7$
Group 7: $A_7 \rightarrow A_8$	Group 8: $A_8 \rightarrow A_8, A_9$
Group 9: $A_9 \rightarrow A_{10}$	Group 10: $A_{10} \rightarrow A_{10} (2), A_{11}$
Group 11: $A_{11} \rightarrow A_{12}$	Group 12: $A_{12} \rightarrow A_{13}, A_{13}$
Group 13: $A_{13} \rightarrow A_{14}$	Group 14: $A_{14} \rightarrow A_{15}$
Group 15: $A_{15} \rightarrow A_{15}, A_{17}, A_{16}$	Group 16: $A_{16} \rightarrow \theta$
Group 17: $A_{17} \rightarrow A_{18}, A_{18}$	Group 18: $A_{18} \rightarrow A_{19}, A_{19}$
Group 19: $A_{19} \rightarrow A_{20}, A_{21}$	Group 20: $A_{20} \rightarrow A_{22}$
Group 21: $A_{21} \rightarrow A_{23}, A_{22}$	Group 22: $A_{22} \rightarrow A_{24}$
Group 23: $A_{23} \rightarrow A_{25}$	Group 24: $A_{24} \rightarrow A_{25}$
Group 25: $A_{25} \rightarrow A_{26}, A_{26}$	Group 26: $A_{26} \rightarrow A_{28}, A_{27}$
Group 27: $A_{27} \rightarrow \theta$	Group 28: $A_{28} \rightarrow A_{29}$
Group 29: $A_{29} \rightarrow A_{30}$	Group 30: $A_{30} \rightarrow \theta$

nearest integer.

Table 6.3.3. Fuzzified production electric

	Forecasted Value of the fuzzy membership
A1	491.774505
A2	491.774505
A3	1393.862396
A4	1632.009299
A5	1831.56375
A6	2087.223884
A7	2248.782768
A8	2353.177843
A9	2709.848068
A10	2793.960658
A11	3178.71736
A12	3364.194305
A13	3574.17403
A14	3808.70806
A15	4009.706818
A16	4255.915177
A17	4429.174874
A18	4663.311673
A19	4970.420928
A20	5235.54034
A21	5313.69142
A22	5583.381938
A23	5832.471383
A24	5832.471383
A25	6143.896952
A26	6563.989253
A27	6813.575862
A28	7082.85575
A29	7503.43
A30	7953.6358

[Step 5] Calculate the forecasting value. In column 1 from table 6.3.3 we start forecasting our data by using states from the previous year. How we do this? Let's see some examples: We forecast the value of February 2016 by looking at the state of February 2015 and we affect the corresponding forecasted value from column C which is (491, 774505). We forecast the value of July 2016 by looking at the state of July 2015 and we affect the corresponding forecasted value from column C which is (3364, 194305). Table 6.3.4 and 6.3.5 below shows the actual and forecasted values of production and consumption electric for period from January 2016 to December 2019 in GWh, the result has been rounded to the

Table 6.3.4. Actual and forecasted values of production electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Proposed Method	2016	435	492	435	492	1355	1394	1799	1832	2268	2353	2759	2794	3312	3364	3877	4010	4373	4429	4865	4970	5370	5314	6133	6144
	2107	466	492	466	492	1459	1394	1938	1832	2437	2353	2960	2794	3570	3574	4166	4010	4689	4663	5194	5236	5712	5832	6604	6564
	2018	523	492	523	492	1576	1632	2088	2087	2636	2710	3195	3179	3837	3809	4480	4429	5036	4970	5582	5583	6163	6144	7112	7083
	2019	545	492	545	492	1675	1632	2220	2249	2781	2794	3367	3364	4030	4010	4703	4663	5297	5314	5874	5832	6480	6564	7503	7503

Table 6.3.5. Actual and forecasted values of consumption electric in GWh

Model	Month	Jan		Feb		Mar		Apr		May		Jun		Jul		Aug		Sep		Oct		Nov		Dec	
	Year	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast	Actual	Forecast
Proposed Method	2016	495	537	876	929	1352	1377	1809	1868	2282	2271	2776	2797	3329	3371	3892	3892	4389	4245	4878	4802	5385	5415	6130	5854
	2107	487	537	936	929	1446	1377	1931	1868	2426	2271	2951	2972	3558	3539	4157	4128	4689	4716	5202	5211	5733	5687	6364	6464
	2018	600	537	1055	929	1588	1621	2109	2191	2663	2635	3229	3178	3878	3892	4530	4245	5106	5059	5655	5687	6220	5854	6900	6885
	2019	617	537	1106	1092	1680	1621	2233	2271	2799	2797	3398	3371	4065	4128	4742	4802	5344	5415	5923	5854	6514	6464	7226	7226

7. Conclusions

The forecasts obtained utilizing Fuzzy time series method and Fuzzy Time Series and Clustering and Proposed Method are discussed in this paper. The aforementioned methods require only the historical data series of electricity consumption to build the forecast. This can be considered as an important advantage, because the effort and cost linked to the data mining are very limited. These historical time series data are analyzed to understand the past and predict the future. Mean Absolute Percentage Error the mean absolute deviation.

The results of predictive metrics of electricity production indicated that the Mean Absolute Percentage Error ranged from 0.95 (Proposed Method) to 30.06 Fuzzy Time Series method and mean absolute deviation ranged from 14.97096316 (Proposed Method) to 231.2533488 Fuzzy Time Series method, Fuzzy Time Series and Clustering method performed closed to the Proposed Method but Fuzzy Time Series method deviated a lot (Table 7.1), (Table 7.2). So, this criterion clearly indicated the superiority of Proposed Method in forecasting the production of electricity and consumption electric during 2016-2019. Similarly, the Proposed Method gave the lowest value thus performed best followed more than Fuzzy Time Series and Fuzzy Time Series and Clustering methods. Fuzzy Time Series method performed the worst in all cases. Similar to the prediction of electric energy production, Proposed Method performed best when the consumption of electric energy was predicted. But it is quite interesting that in case of forecasting the electric energy consumption.

Table 7.1. The values of evaluation metrics production of electricity

Method	MAPE	MAD
Fuzzy Time Series	30.06	231.2533488
Fuzzy Time Series and Clustering	19.54	103.735
Proposed Method	0.95	14.97096316

Table 7.2. The values of evaluation metrics the consumption of electricity

Method	MAPE	MAD
Fuzzy Time Series	26.85	228.420
Fuzzy Time Series and Clustering	13.81	80.26689528
Proposed Method	0.67	16.2021658

Statistical model forecast the future demand and production by a mathematical combination of previous demand and production as well as incorporate other exogenous factors like weather conditions, seasonality etc. The forecasting accuracy depends not only on the numerical efficiency of the algorithms employed, but also on the quality of data analyzed and the ability to incorporate important fundamental factors, such as historical demand, demand and consumption forecasts, weather forecasts of fuel prices. In our study, while all these factors were considered,

there is no doubt that Proposed Method performed the best forecasting. So, this method can be used to create a balance sheet between future forecasting for electric energy production and consumption, According to our study, the future consumption of electric energy will increase at a higher rate in next three years as our prediction showed more consumption of energy than production in most of the months throughout a year (table 7.3 and table 7.4), and balance energy for production and consumption of electric show that some months we need product more electric for To fill the electricity shortage (Figure 4). As Proposed Method produced more reliable information, it can be suggested that it is time, government took necessary steps to produce more electric energy for the future to cope up with the future demand.

Table 7.3. The prediction of Production Electric

Month	2021	2022	2023
Jan	491.989206	503.3697422	510.8906906
Feb	491.989206	503.3697422	510.8906906
Mar	1962.569219	2214.920883	2415.731294
Apr	2710.012313	2711.082843	2716.563877
May	2802.805653	2711.082843	2716.563877
June	3829.126754	4026.063991	4249.983637
July	4440.024397	4713.770666	5000.590303
August	5278.439903	5598.919659	5881.625537
September	5901.678084	6181.533648	6508.320179
October	6511.053799	6946.244241	7284.897621
November	7262.193741	7727.090906	8116.520842
December	8229.341832	8582.642187	8828.813408

Table 7.4. The prediction Consumption of Electric

Month	2021	2022	2023
Jan	547.0917457	557.4954711	571.6345653
Feb	1118.555133	1031.676219	1092.273382
Mar	1632.042481	1659.288374	1630.939523
Apr	2365.020587	2502.590615	2643.732261
May	3286.816771	3400.675637	3586.340934
June	3857.29854	4083.783012	4276.190583
July	4549.21054	4773.14994	5066.664609
August	5526.750832	5905.615987	6272.6894
September	6042.666941	6361.358987	6685.357146
October	6835.374338	7255.662489	7752.835402
November	7544.952606	8031.050891	8214.429822
December	8005.489112	8410.589643	8699.081182

This paper compared several methods to forecast production and consumption electric energy for the period of 2016 to 2019. This study has fulfilled the objectives of the study to propose the production and consumption of future electricity by different forecasting methods like Fuzzy Time Series method and Fuzzy Time Series and Clustering method, and Proposed Method, and then inspect the accuracy of both

models in forecasting ability.

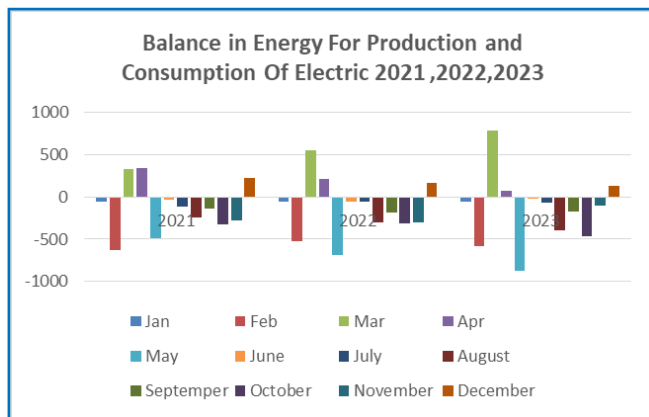


Figure 4. Balance in Energy Production and Consumption for Electric

After examining several models, it was found that Proposed Method was the best and most appropriate to apply to study data series of production and consumption electricity. By analyzing the forecasted value using the performance evaluation procedure, it is found that use of Proposed Method for forecasting production and consumption electricity is better Fuzzy Time Series method and Fuzzy Time Series and Clustering method.

In the end, we can conclude that Proposed Method forecasting model help to determine the full intervals of lengths in the time series and it provides proper intervals according to different observations. On the other side the second model, the Fuzzy Time Series and Clustering method focuses on logical principles that utilize in cluster multi-dimensional data and it assigns the Centre of the cluster from 0 to 100 per cent according to the requirement of the cluster. As compared to traditional hard threshold clustering, it is considering more powerful and every point is label exactly. Clustering of fuzzy time series can be accomplished through proper applications and according to the classification of new kind of data the output of this function can be repurposed.

These three methods and their studies help to explain that forecasting production is very important according to production scenario and in this regard, the data sets are utile and measures according to the nature of production. Consumption of electric and forecasting production plays a major role and for these developments, these two models show their major working according to their data sets and how to manage the data as per their required position, algorithms are used according to nature and utilization of datasets.

REFERENCES

- [1] Boßmann, T., Schurk, R., & Schleich, J. (2015). Unravelling load patterns of residential end-uses from smart meter data. *Proceedings of the ECEEE Summer Study on Energy Efficiency*, 1(6).
- [2] Sadaei, H. J., e Silva, P. C. D. L., Guimarães, F. G., & Lee, M. H. (2019). Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy*, 175, 365-377.
- [3] Ozawa, K., Niimura, T., & Nakashima, T. (1999, May). Fuzzy time-series model of electric power consumption. In *Engineering Solutions for the Next Millennium. 1999 IEEE Canadian Conference on Electrical and Computer Engineering (Cat. No. 99TH8411) (Vol. 2, pp. 1195-1198)*. IEEE.
- [4] Barton, J., Huang, S., Infield, D., Leach, M., Ogunkunle, D., Torriti, J., & Thomson, M. (2013). The evolution of electricity demand and the role for demand side participation, in buildings and transport. *Energy Policy*, 52, 85-102.
- [5] Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1), 68-80.
- [6] Severiano, C. A., Silva, P. C., Sadaei, H. J., & Guimarães, F. G. (2017, July). Very short-term solar forecasting using fuzzy time series. In *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE) (pp. 1-6)*. IEEE.
- [7] Bose, M., & Mali, K. (2019). Designing fuzzy time series forecasting models: a survey. *International Journal of Approximate Reasoning*, 111, 78-99.
- [8] Ellegård, K., & Cooper, M. (2004). Complexity in daily life – a 3D-visualization showing activity patterns in their contexts. *Electronic International Journal of Time Use Research*, 1(1), 37-59.
- [9] Silva, G. C., Silva, J. L., Lisboa, A. C., Vieira, D. A., & Saldanha, R. R. (2017, November). Advanced fuzzy time series applied to short term load forecasting. In 2017.
- [10] Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1-37.
- [11] Wu, W.-Z., Xu, Y.-H.: Rough Approximations of Intuitionistic Fuzzy Sets in Crisp Approximation Spaces. In: *Proceedings of Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2010)*, vol. 1, pp. 309-313 (2010).
- [12] L. A. Zadeh, "Fuzzy sets," *Information and Computation*, vol. 8, pp. 338-353, 1965.
- [13] Rosenfeld, "Fuzzy groups," *Journal of Mathematical Analysis and Applications*, vol. 35, pp. 512-517, 1971.
- [14] *Introduction to the Theory of Fuzzy Subsets-Vol. 1: Kaufmann* (New York: Academic, 1975, 432 pp.).
- [15] Fuzzy Sets and Systems 28, 313-331 (1988) Epstein, B.: Some applications of the... 199-249 (1975) Zimmerman, H.J.: Fuzzy set theory and its applications.
- [16] D. Dubois and H. Prade, *Fuzzy Sets and Systems*.
- [17] Aladag, C. H., Yolcu, U., Egrioglu, E., & Dalar, A. Z. (2012). A new time invariant fuzzy time series forecasting method based on particle swarm optimization. *Applied Soft Computing*, 12(10), 3291-3299.
- [18] Q. Song and B. S. Chissom, *Forecasting enrolments with fuzzy time series —Part I, Fuzzy Sets and Systems* 54 (1993)

- 1–9.
- [19] Q. Song and B. S. Chissom, Forecasting enrolments with fuzzy time series —Part II, *Fuzzy Sets and Systems* 62 (1994) 1–8.
 - [20] Ellegård, K., & Palm, J. (2015). Who Is behaving? Consequences for energy policy of concept confusion. *Energies*, 8(8), 7618–7637.
 - [21] Z. Ismail and R. Efendi, Enrolment forecasting based on modified weight fuzzy time series, *Journal Artificial Intelligence* 4(2011) 110–118.
 - [22] Ellegård, K., & Vrotsou, K. (2006). Capturing patterns of everyday life: Presentation of the visualization method VISUAL-TimePacTS. Paper presented at the International Association for Time-Use Research Conference, Copenhagen, Denmark.
 - [23] H. K. Yu, Weighted fuzzy time series models for TAIEX forecasting, *Physica A* 349 (2005) 609–624.
 - [24] Ellegård, K., Vrotsou, K., & Widén, J. (2010). VISUAL-TimePacTS/energy use: A software application for visualizing energy use from activities performed. Paper presented at the 3rd International Scientific Conference on 'Energy systems with IT'.
 - [25] Finn, P., O'Connell, M., & Fitzpatrick, C. (2013). Demand side management of a domestic dishwasher: Wind energy gains, financial savings and peak-time load reduction. *Applied Energy*, 101, 678–685.
 - [26] Gaiser, K., & Stroeve, P. (2014). The impact of scheduling appliances and rate structure on bill savings for net-zero energy communities: Application to West Village. *Applied Energy*, 113, 1586–1595.
 - [27] Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
 - [28] Estivill-Castro, Vladimir (20 June 2002). "Why so many clustering algorithms – A Position Paper". *ACM SIGKDD Explorations Newsletter*. 4 (1): 65–75 (11).
 - [29] M. S. Aldenderfer and R. K. Blashfield. *Cluster Analysis*. Sage Publications, Los Angeles, 1985.
 - [30] M. R. Anderberg. *Cluster Analysis for Applications*. Academic Press, New York, December 1973.
 - [31] Anderson, B. (2017). Laundry, energy and time: Insights from 20 years of time-use diary data in the United Kingdom. *Energy Research & Social Science*, 22, 125–136.
 - [32] Ucal and B. Oztaysi, Forecasting energy demand using fuzzy seasonal time series, in *Computational Intelligence Systems in Industrial Engineering*, Atlantis Computational Intelligence System, Vol. 6 (Atlantis Press, 2012).
 - [33] EUROSTAT. (2009). *Harmonised European time use surveys 2008 guidelines*. Luxembourg: Office for Official Publications of the European Communities.
 - [34] Flath, C., Nicolay, D., Conte, T., van Dinther, C., & Filipova-Neumann, L. (2012). Cluster analysis of smart metering data. *Business & Information Systems Engineering*, 4(1), 31–39.
 - [35] Pan, S., Wang, X., Wei, Y., Zhang, X., Gal, C., Ren, G.,... & Xie, J. (2017, December). Cluster analysis for occupant-behavior based electricity load patterns in buildings: A case study in Shanghai residences. In *Building Simulation* (Vol. 10, No. 6, pp. 889–898). Tsinghua University Press.
 - [36] M. Sah and K. Y. Degtiarev, Forecasting enrolment model based on First-order fuzzy time series, *Proceeding of World Academy of Science, Engineering and Technology* 1(2005) 375–379.
 - [37] M. Sah and K. Y. Degtiarev, Forecasting enrolment model based on First-order fuzzy time series, *Proceeding of World Academy of Science, Engineering and Technology* 1(2005) 375–379.
 - [38] S. M. Chen and C. C. Hsu, A new method to forecast enrolments using fuzzy time series, *Applied Science and Engineering* 2(2004) 234–244.
 - [39] H. S. Lee and M. T. Chou, Fuzzy forecast based on fuzzy time series, *International Journal of Computer Mathematics* 81 (2004) 781–789.
 - [40] H. T. Liu, An integrated fuzzy time series forecasting system, *Expert Systems with Applications* 36 (2009) 10045–10053.
 - [41] Sutton, D. C. (2003). Cluster analysis. *International Journal of Market Research*, 45(4), 515–518.
 - [42] Z. Ismail and R. Efendi, Enrolment forecasting based on modified weight fuzzy time series, *Journal Artificial Intelligence* 4(2011) 110–118.
 - [43] Palm, J., Ellegård, K., & Hellgren, M. (2018). A cluster analysis of energy-consuming activities in everyday life. *Building Research & Information*, 46(1), 99–113.