REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

APPLICATION OF CHAOTIC TIME SERIES FOR CASTING CONSUMER DEMAND

Speciality:	3338.01 – “System anlysis, control and information processing “(management and decision-making)"
Field of science:	Engineering sciences
Applicant:	**Sevinj Rasul Mustafayeva**

Baku -2022

The work was performed at Azerbaijan State Oil and İndustry University, Department of Computer Engineering

Scientific supervisor : doctor of engineering sciences, professor
Natig Hajı Javadov

Rəsmi opponentlər: doctor of engineering sciences, professor
Mahammad Nurmahammad Nuriyev

doctor of engineering sciences, professor
Salahaddin İmamali Yusifov

doctor of engineering sciences, professor
Javanshir Firudin Mammadov

Dissertation council ED 2.02 of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and İndustry University

Chairman of the
Dissertation council

_______________ doctor of engineering sciences, professor
Mustafa Baba Babanli

Scientific secretary of
the Dissertation council

_______________ doctor of phylosofy in engineering associate, professor
Tahir Gaffar Jabbarov

Chairman of the
scientific seminar

_______________ doctor of engineering sciences, professor
Tarlan Samed Abdullayev

GENERAL DESCRIPTION OF THE DISSERTATION

Relevance of the topic. The problem of analysis and processing of time series in the context of principles of synergetic have special actuality.

Time series are the main result of both natural and calculational experiments. The tasks of meteorology, geophysics, and financial analysis are usually indicated as priorities. In recent times, physiology, medicine, and social sciences have joined. There are several types of tasks usually solved in the processing of time series. Currently, the problem of analyzing and predicting a high number of heterogeneous interacting information flows in complex network structures continues to be relevant. Evolutionary trajectories in the state space of the system become sensitive to small information influences (fluctuations). As a result of the cumulative effect, at a certain moment the system can move from one trajectory of evolutionary development to another, and the system's behavior is not defined by the energy of influences, but its parameters and form.

During the evolution process in an open system, the growth of information flows leads to a complication of the information component, that leads to an increase of chaotic processes, which transforms the system into a state of dynamic chaos. Additionally, the heterogeneous component of information flows has a significant impact on the coherence processes (in the wide sense), which decreases the opportunities for efficient functioning of the system.

The following major tasks are highlighted in the applied analysis of time series: obtaining, the identification and prediction task.

Since predictions represent a wide range of thoughts and opinions - as markets in general, prediction markets have been quite effective as a predictive tool. As a result of their foresight, prediction markets (sometimes known as virtual markets) have been used by a number of major companies. The mixing of economics, politics and in recent times, cultural factors have only increased the demand for forecasting (prediction). Add the benefits of data analytics and artificial intelligence; we live in a golden age of data and statistical usefulness.

However, traditional methods of realizing these tasks require a significant amount of information or stationary data series, which cannot be always obtained in practice during the investigation of real systems. Therefore, the problem of identification and prediction of time series, in the context of the principles of synergetic, becomes particularly actual.

Research object and topic . Determination of consumer demand of automotive object based on a fuzzy logic under uncertainty, analysis of the relationship between fractality and fuzzy model.

Goal and tasks of the dissertation. The goal of the dissertation is to provide a prediction of consumer demand on the basis of chaotic time data in the automobile field based on previous years' sales data.

For achieving the goals, the following tasks were assigned and solved by the author:

- presentation of car sales data in the form of a time series;
- determination of the positive Lyapunov index of time series describing car sales;
- determination of the chaotic state of the time series data;
- fractal analysis of chaotic time series and identification of fractal dimensionality;
- Analysis of the long-term memory effect of chaotic time series;
- identification of the connection between the fractal dimensionality and the number of rules of fuzzy logic models;
- prediction of consumers' demand for automobile sales using If-Then models.

Research methods. Synergetic principles, nonlinear dynamics, chaos theory, control theory of chaotic process, functional analysis, methods of fractal analysis, methods of fuzzy logic were used as research methods in the dissertation.

Computer simulation in MATLAB environment was conducted and the accuracy of the received results was confirmed.

Main highlights, brought forward for dissertation defense.

In the dissertation, the following issues(points) were reviewed:

- Development of a mathematical model of the system, considering a new reading of chaos.
- Development of an algorithm for non-parametric forecasting based on the calculation of fractal dimensionality and the possibilities of fuzzy logic.
- Conducting a cluster analysis of chaotic time series, considering the connection of fractality with the number of rules;
- Considering the " If-Then" rule of the fuzzy model demand prediction using chaotic time series.

Scientific novelty . The main scientific innovations obtained in the dissertation are claimed as folows:

- have been conducted a multidimensional study on the identification of the time series on the sales of several years.
- has been analyzed the dynamic behavior of the system in terms of a new reading of the chaos theory.
- have been determined the fractality of chaotic time series and was proved the effect of long-term memory.
- established the connection between the fractal dimensionality and the number of fuzzy model rules
- the fuzzy stability of financial time series characterized by uncertainty has been studied for the first time.
- considering the "If-Then" rule of fuzzy model, the consumer demand of financial chaotic time series is predicted.

Theoretical and practical significance. The theoretical significance lies in the fact that the development of the system "observables-identification-control" and the inclusion of invariant sets in the synergetic methods of analysis and control of chaotic and systems make a weighty contribution, supplementing the existing traditional methods of analysis of observables in the space of control.

Usage of the fuzzy logic method makes it possible not to impose special requirements on the data and allows to visually investigate the behavior of the observables and to make predictions.

The scientific results received in the dissertation might be applied to the marketing in various spheres. The main practical significance lies in the possibility of determining the demand, which in turn allows budget planning and cost reduction.

Approbation of the dissertation. The results obtained in the dissertation were heard at the following conferences:

1. Мустафаева С.Р. Графические методы идентификации и прогнозирования взаимодействующих временных последовательностей // Международная научно–практическая конференция молодых ученых и студентов «Информационные процессы и технологии, - Севастополь: Изд-во СевНТУ, - 23-27 апреля, - 2012, стр.267.
2. Мустафаева С.Р. Анализ взаимодействующих хаотических временных рядов //Третья Международная научно-практическая конференция «Информационные технологии и компьютерная инженерия», - Винница: Изд-во ВНТУ 29-31 мая 2012 стр.262-263.
3. Мустафаева С.Р. Anomalous diffusion influence of chaotic system by observation // IV International conference “Problems of cybernetics and informatics”, - Baku, - September 12-14 ,-2012, pp.158-160.
4. Мустафаева С.Р. Идентификация и прогнозирование временных рядов // Doktorantların və gənc tədqiqatçıların Azərbaycan xalqının Ümummilli lideri Heydər Əliyevin anadan olmasının 90 iliyinə həsr olunmuş” Azərbaycan 2020: neft-qaz sənayesinin inkişaf perspektivləri” konfrans, - Bakı: - 2-3 may,- 2013, s.183-185.
5. Мустафаева С.Р. Нелинейные методы обработки временных рядов //Международная научно-практическая конференция молодых ученых, аспирантов и студентов, - Харьков: Изд-во, - 2013, стр.82.
6. Mustafayeva S.R. Calculation of Lyapunov exponent of time series //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-2019 Switzerland, pp.768-771.

7. Mustafayeva S.R. Fractal Dimension Determining for Demand Forecasting //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-Switzerland, pp.931-937.

8. Мустафаева С.Р. Использование метода кластеризации в автомобильном бизнесе, ADNSU-nun 100 illiyinə həsr olunmuş ADNSU gənc alim və tədqiqatçıların elmi-praktiki konfransı 2020- Bakı: - 2020 ADNSU, стр.452-457.

Publications. 18 scientific works were published on the subject of the dissertation: 8 articles were published in peer-reviewed scientific journals, and 10 were published in collections of conference materials. The main indexation of the listed works is reported as follows:

2 articles were published in the collection of conference proceedings indexed in Web of Science.

2 articles were recommended by AR AAK.

6 articles were indexed by "РИНЦ" and published in journals recommended by RF AAK.

Name of the organization within which the dissertation work is carried out. Dissertation work is carried out at the "Computer Engineering" department of Azerbaijan State Oil and Industry University.

Composition and volume of the dissertation. The dissertation consists of an introduction, 4 chapters, conclusions, a list of 150 references and appendices. The work is written on 155 pages, including 3 tables and 13 pictures. The total volume of the dissertation is 175,000 characters.

DISSERTATION S CONTENTS

The introduction presents the actuality of the work, scientific novelty, research methods, the main defended provisions and the practical significance of the dissertation.

In the first chapter describes the status, analysis of the problems and opportunities of time series in the various tasks implementation of marketing and their activity program. The goal and tasks necessary to carry out the work are formulated.

Sales are a vital part of any business. Accordingly, sales prediction plays an important role in the business planning and is a company self-assessment tool. Managers must keep their hand on the pulse of their company. Sales prediction reports, graphs are analyzing the pulse of the business. This can make the difference between simply surviving and being highly successful in business. It is an important cornerstone of a company's budget. In the future, a company's direction can depend on the accuracy of its sales predictions. Prediction allows companies to effectively allocate resources for future growth and manage cash flow.

One of the first achievements in this area is the work of White (1988), as well as other related articles that use sophisticated networks: Bosarge (1993), Wong (1990), Hiemstra (1996) and Hefke and Helmenstein (1996), which find nonlinearities in time series and improve prediction. Other contributions with prediction for other macroeconomic variables are Franses and Draisma (1997) and Swanson and White (1997). Prediction can be done for closing/opening stock prices on a daily basis, quarterly company earnings, etc. Various models are available in the literature for predictions of these time series.

Some of them are:

- Autoregressive Integration Moving Average (ARIMA)
- Simple Moving Average (SMA)
- Exponential smoothing (SES)
- Neural Network (NN)
- Linear Regression Models
- Logistic Regression
- Reference Vector Machines
- Naive Bayesian
- Hidden Markov
- VAR
- Gaussian processes.

Requires methods that are relatively simple to implement and easy to interpret. Ideally, the results of the preliminary analysis should indicate directions for further analysis, and then further directions for research. Synthesis of new analysis methods in a consistent, unified form will allow researchers to conduct analysis more qualitatively and confidently. The time series approach is most effective for well-established enterprises with several years of data on which to work alongside relatively stable trend models.

On the basis of the above analysis in the field of car sales, we can once again emphasize the necessity of research in predicting consumer demand. The model used in this study does not require special knowledge about the data collection used. The advantage of using fuzzy logic is to handle imprecise, uncertain, fuzzy, semi-true or approximate and non-linear data.Considering the above-mentioned shortcomings, in the dissertation the task statement is formed.

The goal of the research is the prediction of consumer demand using chaotic time series based on fuzzy logic. To achieve this goal, it is necessary to solve the following problems:

- to develop a mathematical model of the system considering a new reading of chaos.
- to develop an algorithm for non-parametric prediction based on the calculation of fractal dimensionality and the possibilities of fuzzy logic.
- to conduct a cluster analysis of chaotic time series, considering the connection of fractality with the number of rules;
- considering the "If-Then" rule of fuzzy model predicted demand using chaotic time series.

Second chapter is devoted to the development of an intelligent adaptive system "observables-identity-management".

In general, the time series modeling task is known as "systems identification" and "systems prediction".

In the dissertation, the state and prospects of implementation of the problem under study will be considered in the context of the identification task and the prediction task.

The identification task. When solving it the attempt is made to answer the question, what are the parameters of the system, which generated a given time series. Parameters can be very different - statistical distributions, statistical model parameters, spectral properties, etc. It is important that these parameters would help to identify the system (process), that is, to distinguish it from others.

The task of prediction. It consists in predicting the future values of the measured characteristics ,or more widely, the future state of the analyzed object on the basis of observational data. If small perturbations of the initial condition do not increase over time (that is, there is stability), then the behavior of such a system is predictable. Otherwise, the process can only be described in a probabilistic way. It is these considerations that formed the basis of the modern idea of dynamic chaos.

The importance of studying chaos is that chaos offers an alternative method to explain the apparently random behavior of complex systems. Chaos plus specific mathematical tools is the basis for studying various models from different areas, which can be reduced to elementary ones with known chaotic behavior for some parameter values.

In the common case, a scalar time series (TS) will be called an array of numbers representing the values of some dynamic variable with a constant step in time, i.e., at the moments.

$$t_i = t_0 + (i-1)\Delta t\,;\; x_i = x(t_i)\,,\; i = \overline{1, N}\,. \qquad (1)$$

Time series are data collections that represent the performance of one (or more) random variables over time, and their main characteristic is that sequential entries of that variable, which not independent of each other, and their analysis must consider the sequence in which they have been collected.

Calculation of Lyapunov indicators by realization. One of the features of chaotic regimes is the instability of each trajectory belonging to a chaotic attractor.

Lyapunov's characteristic indicators have proved to be a quantitative measure of this instability. They are one of the most important characteristics of an attractor, since they allow estimating:

- fractal dimension of the attractor;

- entropy of the dynamic system;
- the characteristic time of predictability of the system's performance.

In the process of evolution in an open system, the growth of interacting information flows and objects leads to the complication of the information component, which in turn causes an increase in chaotic processes that transfer the system into a state of dynamic chaos.

Since the criterion of chaotic dynamics is the presence of a positive senior Lyapunov index, it is of great interest to estimate it on the basis of processing time series (realizations).

x (t)- be the typical phase trajectory of the system x=v (x, a) and x_1(t) a trajectory close to it, $x_1(t)=x(t)+\xi(t)$ [14, pp.419-431]. Now consider the function:

$$\Xi(\xi(0))=\lim_{t\to\infty}\frac{1}{t}\ln\frac{|\xi(t)|}{[\xi(0)[}, \qquad (2)$$

which is defined on initial displacement $\xi(0)$ vectors such that $|\xi(0)|=\varepsilon$, where $\varepsilon\to 0$. Depending on the direction of the vector $\xi(0)$, a function $\Xi(\xi(0))$ will then take a final series of values , $\{\lambda_i\}$, $i=1,2,...,n$, , which are called Lyapunov characteristic indexes.

Lyapunov's characteristic indexes serve as a measure of the randomness of dynamical systems. In particular, if positive indicators are available, then the system conducts in a chaotic manner.

Now consider a time series obtained by recording the observed variable at successive equally spaced points in time $x_1, x_2, ..., x_N$.

Setting some integer *m* and using the idea, the reconstruction of the attractor by the method of lagging in the space of dimension *m*, that is, put

$$x_i=(x_1, x_{i-p}, x_{i-2p}, ..., x_{i-(m-1)p}),\ i=1,2,... \qquad (3)$$

Considering the multitude of points x_i in the *m*-dimensional space as a phase portrait of the attractor, estimate its dimensionality by the Grasberger-Procaccia method, that is, calculate the correlation integral for different ε

$$\tilde{N}(\varepsilon) = \lim \frac{1}{M(M-1)} \sum_{i,j=1}^{M} \theta\left(\varepsilon - \left\| x_i - x_j \right\|\right) \qquad (4)$$

plot the obtained dependence on the graph in coordinates ($log\, \varepsilon, log\, C(\varepsilon)$), search on it a linear segment, and determine the angular coefficient D*(m)*. Perform the described procedure several times, considering sequentially $m = 1,2,3,...$

Here it is important to note that the presence or absence of dependence D*(m)*. saturation during magnification *m* is considered as a criterion of whether the signal is generated by a dynamical system or is noise. If saturation is observed at some level *D*, then value *D* is taken as an estimate of the correlation dimensionality of the DS attractor that generated the observed signal. Based on the estimation *D* and the Mane theorem , we may conclude that the dimensionality of the phase space of this D does not exceed *2D+1*.

The statement proved by Mane is that any practically smooth mapping *X* in *Y* will specify an embedding of a multitude *A* in space *Y,* under the condition that

$$m \geq 2D_A + 1, \qquad (5)$$

where D_A - fractal dimension of the multitude *A*; *m*- embedding dimension.

Fuzzy logic is based on fuzzy set theory, which is a summary of classical set theory (Zadeh, 1965). The problem of identifying fuzzy systems is the problem of extracting IF-THEN rules from raw input and output data. This occurs as follows:

1) Clustering.

2) Specification of input-output relations (IF-THEN rules).

The third chapter is devoted to the intellectual analysis and identification of the characteristics of chaotic time series that describe the dynamics of car sales.

As mentioned earlier, given the transparency and dynamism of financial processes, the system is described by time series. The data represent car sales for 2012-2013.

Table 1.
Sales data of car sales by official dealer for 2 years

2012	2013
56	46
54	65
67	75
111	70
99	78
63	80
133	88
79	91
87	93
71	49
86	65
135	63

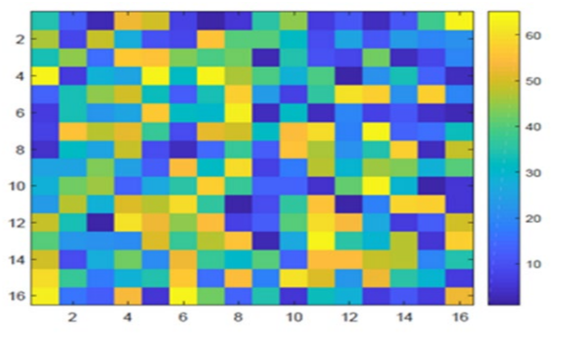

Figure 1. Chaotic time series fragment for 2 years of car sales

In order to prove the chaotic nature of our time series we determine the Lyapunov index.

1. Enter the data for the sales of 2 years.
2. Using the program in MATLAB find the value of the Lyapunov index

function zout = zlap (I, J, X, TAU)

% m file for function zlap used in lap.m for calculation of

% the first Lyapunov exponent using the method described by Wolf

zout = X(I+(J-1) *TAU);

Lyapunov exponent

Lamda1=0.6609

Using MATLAB software package determine Lyapunov index L= 0.6609.

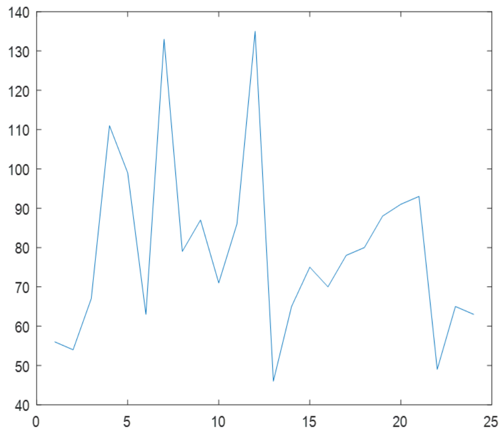

Figure 2. Graphical representation of the Lyapunov index

The positive Lyapunov index shows the existence of chaos in a dynamical system with bounded trajectories. Then, according to the algorithm, we should determine the fractal dimension of this system. First of all, we need to disclose the concept of fractal analysis of time series.

Before proceeding to the definition of fractal dimension, we shall introduce the concept of "fractal". Fractal can be defined as follows:

"A curve or geometric figure, each part of which has the same statistical character as a whole. Fractals are useful in modeling structures (such as fuzzy shorelines or snowflakes) where similar patterns repeat at progressively smaller scales, and in describing partially random or chaotic phenomena such as crystal growth, fluid turbulence and galaxy formation."

There are many methods to evaluate quantitatively the fractal characteristics of a structure through fractal dimensionality.As a traditional example, the fractal dimensionality of a spatial fractal structure can be determined quantitatively using a box-count fractal analysis, which examines the method by which the structure fills the space. However, such spatial analysis is generally poorly suited to the analysis of so-called "time series" fractals, which may exhibit exact or statistical self-affinity, but which inherently lack well-defined spatial characteristics.

Fractal is a geometric object with internal homothety (from the Greek Ομοιο-θετικός, similar shape), which means that it repeats its shape in the same form at different scales. Magnifying any part, it is possible to obtain a number that is similar to the original. Therefore, it is called fractal geometry. Fractal geometry comes down to identifying these configurations, analyzing and manipulating them. Besides being a tool for analysis and synthesis, it can also be used as a prediction tool. The rules of fractals are precise, thus the outcome is predictable. This is in contrast to traditional science, which instead looks at irregular aspects of nature and non-similar events.

The determination of fractal dimensions of systems is well known to enable predicting their performance. Therefore, in this research were conducted a study in the context of selecting the area of interest, analyzing and determining its fractal dimensionality. In this case we want to determine the fractal dimension of the time series to predict the market demand in the next year according to the available data for the last years. In particular, such characteristic of time series as fractal dimension allows to determine the moment, when the system becomes unstable and ready for transition to a new state.

The fractal dimensionality of the time series is determined using MATLAB Simulink. At first, we convert our data from a linear form to a three-dimensional form. Then we determine the fractal dimensionality. Consequently, a time series has a long-term memory effect, assessing its depth, we reveal the presence of cycles (quasi-cycles). And we can say that our result characterizes investments in the corresponding financial instrument as "relatively risk-free".

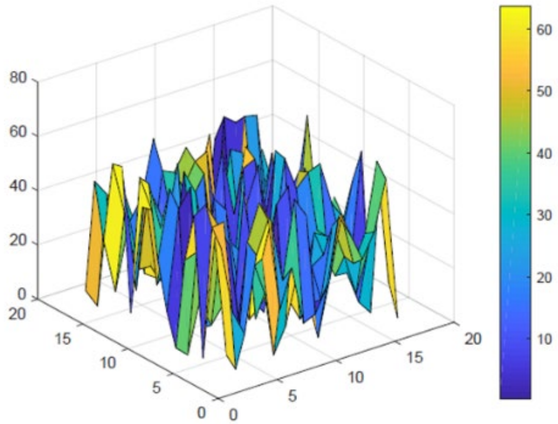

Figure 3. Fractal dimensionality obtained in MATLAB

Traditionally, we assign to a particular geometric object a precise value of a fractal and this numerical value has been considered to be a specific label for the object. Fractal analysis has been useful in describing dynamic phenomena. It has also lead to a number of applications. This term refers to the characterization of the global property of self-similarity, which describes how much a part of a phenomenon can be enlarged to the whole. Instrument used for a quantitative assessment of this effect is the fractal dimensionality. For example, in time series, fractal dimensionality indicates the regularity of the series and indicates how much the similarity increases when such time series are observed over a longer time interval. Self-similarity may also be viewed as a measure of the geometric complexity of the object under discussion. The definition of fractal dimensionality is inseparably connected with set-based constructions.

Common to the existing methods (despite the obvious technical differences) is that all of them use sets regarded as information pellets, which allow us to see only a certain part of the phenomenon. Changing the size of the information pellets means that how much more of the part we are considering.

Granular calculations are the cornerstone of processing in intelligent systems. The human perceives the word, systematizes knowledge and makes it highly productive through the formation and manipulation of information granules. Information granulation is an example of abstraction. There are many aspects of granular processing of information, and there are many formal structures in which such granulation of information is performed. For example, they include set theory, fuzzy sets, random sets, rough sets, and many others.

It is becoming quite obvious that there may be some interesting and potentially useful connections between the fundamental concepts of fractals and granular calculations. Fractals use information granules to determine their main characteristics.

In this study (research), our goal is to identify the connections between fractal analysis and the roles of information granules played in it. In particular, we review the main idea of calculating fractal dimensionality with fuzzy sets.

In this case it is usual to use correlation measurement, where we build hyperspheres around separate points in the data set, see Figure 4. Following this way, the time complexity depends on the number of data items. Following this method, we calculate the number of data points in a sphere of radius s.

In terms of fractal dimensionality, the complexity of an object is measured, which might be from nature (for example, a bacteria) or a mathematical object created using a formula or algorithm (for example, well known as the Mandelbrot set). However, such numerical values are difficult to use in classification or recognition applications since the calculated values do not correspond to these exact values.

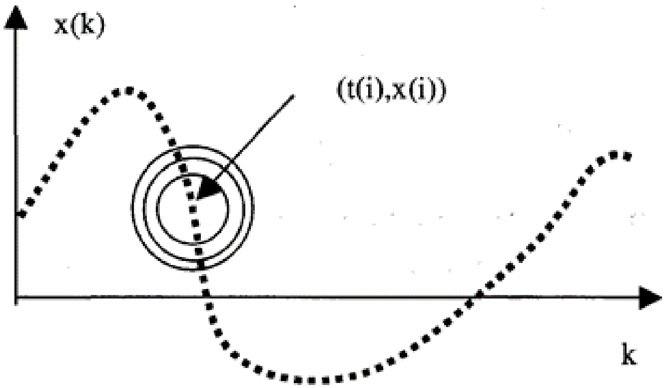

Figure 4. Calculation of Correlation Fractal Dimension

On this basis, the following scheme for estimating the fuzzy fractal dimensionality of a set of geometric objects was proposed. At first, the standard fractal dimensionality of objects is calculated using a box-counting algorithm with point samples from objects.

Secondly, clear values of fractal object dimensions are used to construct linguistic values for measurements, specifically fuzzy fractal object dimensions. As an example, consider a Gaussian function of belonging:

$$A_{ij}(e) = \exp\left(-(x_i - x(f))^2/e^2\right) \tag{6}$$

Assuming that x_i is fixed, the affiliation function returns the affiliation degree of x_j to this vicinity. In this sense, it helps to differentiate (discriminate) the different elements of the time series falling into a given window $Q_{ij}(s)$. As the degree of affiliation A for a given x_j is greater, its contribution to the total sum N(s) is more prominent. The affiliation function itself is equipped with the spreading parameter (e), which controls the size of the information granule. It should be emphasized that fuzzy sets in the above-mentioned definition are defined in the amplitude space, thus the time variable, over which all discrete time moments are distributed, is not affected, and the granulation there is related to the window .

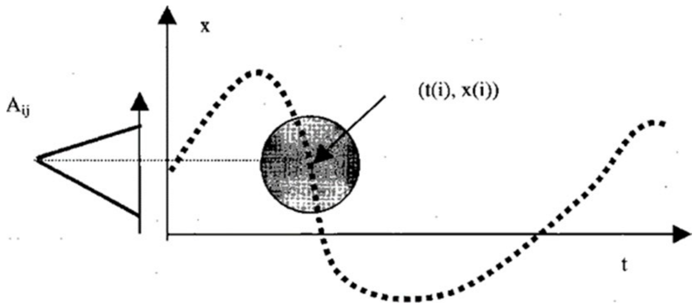

Figure 5. Calculation of fractal dimensionality using correlation dimensionality within fuzzy sets used as information granules

Further, using these linguistic values of fractal dimensions, a set of fuzzy rules is built, so that every object is associated with each rule. Therefore, this set of "if-then" fuzzy rules can be viewed as a classification scheme for a set of geometric objects and can be used to recognize these objects because a particular value is displayed on the object. This is the connection between fractal dimensionality and the number of information granules that are needed for usage in the modeling activity. It is intuitively clear that the more complex the phenomenon being modeled, the more information granules must be used to build the model. The conjuncture we are doing is that the fractal dimension determines the number of fuzzy sets, and the ratio is a power law.

Calculations of local fractal dimensionality help us to identify areas in the data space where we need more "detailed" information granules, and hence the number of rules required to model the data there becomes greater. The step law of detail information becomes useful in distributing the number of fuzzy sets over the regions of the data.

The fourth chapter is devoted to a computer simulation based on cluster analysis (c-means) for identifying consumer demand in the automobile business. The main result of the work is the conclusion about the possibility of predictive analysis of financial time series by calculating the fractal measure. At the same time, this analysis is not an exhaustive tool for the predictive study of chaotic time series, since it does not always provide complete information about their behavior without the use of additional methods and algorithms. Therefore, as mentioned earlier, we will further use the possibilities of fuzzy logic in the MATLAB environment.

Cluster analysis is used for 4 main purposes. These goals are as follows:

a) Divide the n objects into as many different subgroups as possible, each with its own characteristics defined by the variable p;

b) divide the variable p into subgroups that are considered to have common features in terms of values defined in units of n, and identify patterns of common factors;

c) Separate subsets of the unit, p, together with units and variables together;

d) Determine the biological and typological classification followed by the units for the meanings defined.

Stages of cluster analysis:

a) Definition of dimensions and variables used to determine the similarity between units;

b) Multiplying units after finding similarities between units;

c) determining the compatibility of the created clusters;

d) the statistical validity of the imagination of the majority of the population.

This algorithm works by assigning membership to each data point corresponding to each cluster center based on the distance between the cluster center and the data point. As the data is closer to the cluster center, so is affiliation larger to a particular cluster center. Obviously, the sum of the affiliation of each data point must be equal to unity. After each iteration, the affiliation and cluster centers are updated using the following formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{(2lm-1)}} \quad (7)$$

$$v_{ij} = \frac{\left(\sum_{i=1}^{n} (\mu_{ij})^m x_i\right)}{\left(\sum_{i=1}^{n} (\mu_{ij})^m\right)}, \quad \forall \mathrm{j} = 1,2,...\mathrm{c} \tag{8}$$

where n - v_j is the number of data points, *vj*- represents the *j* cluster center, *m*- is the fuzzy index, *c*- presents the number of the cluster center, μ_{ij}- presents the membership of the *i* data center in the *j* cluster center. d_{ij} -presents the Euclidean distance between the *i* data and the *j* cluster center [86, pp. 33-37].
The main goal of the fuzzy c-means algorithm is minimization.

$$J(U,V) = \sum_{i=1}^{n} \sum_{j=1}^{c} (\mu_{ij})^m \left\| x_i - v_j \right\|^2 , \tag{9}$$

where $\left\| x_i - v_j \right\|$- the Euclidean distance between these data and the center of the *j* cluster.

Consider that, $X = \{x_1, x_2, x_3 \ldots, x_n\}$ will be a set of data points, and $V = \{v_1, v_2, v_3 \ldots, v_n\}$ will be a set od centers.

1. Select the cluster centers " c" randomly.

2. Calculate the fuzzy membership μ_{ij} using the above-mentioned formula.

3. Calculate the fuzzy centers " v_j " using (1).

4. Repeat steps 2) and 3) until the minimum value of *" J"* or $\left\| U(k+1) - U(k) \right\| < \beta$ ||. || is reached.

where *"k"* - iteration step. " β " - completion criterion between [0, 1]. "«$U = (\mu_{ij}) n \cdot c$ " - fuzzy membership matrix. *"J"* - target function.

In our example, we want to show the possibility of faster and more accurate prediction using the clustering method with fuzzy logic. Our task is to predict the next year's market demand using the available data. For this purpose, we use MATLAB. Clustering of the numerical data is the basis of many algorithms of classification and system modeling. The purpose of clustering is to identify natural groups of data from a large dataset to create a short summary of system performance. Fuzzy c-means (FCM) is a data clustering method in which a data set is grouped into n clusters, whereby each data set in the data set belongs to each cluster to a certain extent. For example, a particular data object that is close to the center of a cluster will have a high degree of affiliation or belonging to that cluster, and another data object that is far from the cluster center will have a lower degree of affiliation or belonging than that cluster.

Algorithm of actions is as follows.

1. Using the fcmdemo command, you can start a graphical interface that allows you to try out various parameter settings for the fuzzy c-mean algorithm and observe their effect on the resulting two-dimensional clustering. You can choose a sample dataset and an optional number of clusters from the drop-down menus on the right, and then click Start to start the fuzzy c-mean string clustering process. The clustering itself is performed by the fcm function.

2. The fcm function takes the dataset and the desirable number of clusters and returns the optimal cluster centers and affiliation estimates for each data point. You can use this information for building a fuzzy inference system by creating affiliation functions that represent the fuzzy qualities of each cluster. After this operation, we get the following figure.

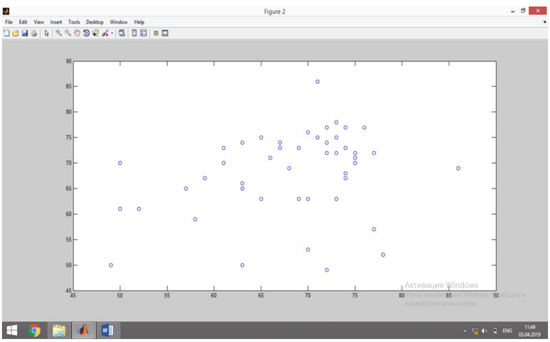

Figure 6. Cluster data of time series over 2 years

Call the fcm command line function in order to find two clusters in this dataset, until the target function is no longer greatly reduced.

[center, U, objFcn] = fcm (fcm data, 2);

Here the central variable contains the coordinates of the two cluster centers, U contains the affiliation estimates for each of the data points, and objFcn provides the history of the target function by iteration.

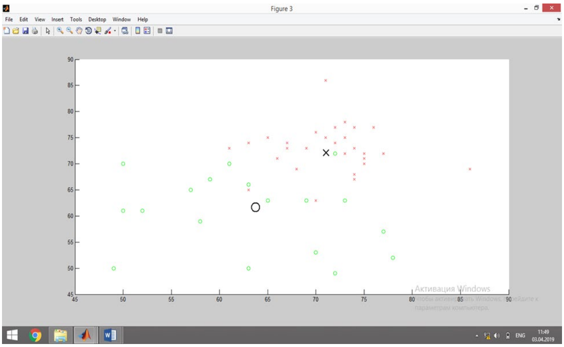

Figure 7. Determination of the center of the clustering target function

The fcm function is an iterative cycle based on the following procedures:

initfcm - initializes the problem

distfcm - calculates Euclidean distance

stepfcm - performs one clustering iteration

In order to see the progress of the clustering, construct the target function.

The verification of the quality of training (learning) reflects the ability of networks to identify hidden patterns in the changes of financial indicators and, as a consequence, the ability to adequately predict the values of financial stability coefficients. Usage and analysis of fuzzy time series and clustering method, extends the capabilities of decision-making systems, thus, provides greater accuracy than traditional methods. Verification of the quality of training (learning) reflects the ability of networks to identify hidden patterns in the changes of financial indicators and, as a consequence, the ability to adequately predict the values of financial stability ratios.

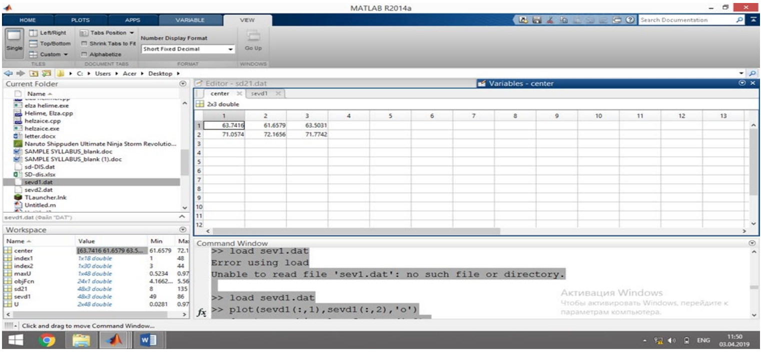

Figure 8. Data obtained from clustering

Usage and analysis of fuzzy time series using the clustering method expands the capabilities of decision-making systems and thus provides greater accuracy than traditional methods.

The prediction task has the purpose of predicting future values of measurable characteristics of the object under study, i.e., to make a prediction for a certain period of time based on observational data. Nowadays, several different prediction methods have been developed and substantiated.

A prediction market is a market in which people can trade contracts that are paid based on the outcome of unknown future events. The market prices generated by these contracts can be seen as a kind of collective prediction by market participants. These prices are based on individual expectations and investors' readiness to risk their money for these expectations.Predictive analytics is often discussed in the context of big data. Technical data, for example, comes from sensors, tools and connected systems around the world. Business system data in a company can include transaction data, sales results, customer complaints and marketing information.

Increasingly, companies are making decisions based on data on the basis of this valuable information.

Successful corporate governance depends on effective strategic and operational planning. Mistakes in planning often lead to enormous costs and, in some cases, to reputational losses. Reliable predictions provide an important contribution to effective planning.

Predicting demand is a very important factor for proper business operations. It is essential for a company to ensure effective operations management planning as all organizations will deal with uncertainty in the future, and some variation should be expected between the forecast and the actual demand. The goal of an accurate demand prediction is to minimize the deviation between actual demand and the forecast. Therefore, due to the uncertainty of demand from customers, they often face obstacles in order to overcome the size and the smallest amount of goods ordered. If a shortage occurs, it could be damaging to the company in terms of costs, while excess inventory could have an adverse effect on further investment. This proves the importance of predicting of consumer demand. The method of fuzzy c-means models is more appropriate due to the use of precise and imprecise data in demand analysis. The centers of clusters obtained by the clustering method are given in the code book below:

```
Center
kl1mfs=u (1, :); kl1mf=kl1mfs';a1=max(kl1mf);
kl2mfs=u (2, :); kl2mf=kl2mfs';a2=max(kl2mf);
kl3mfs=u (3, :); kl3mf=kl3mfs';a3=max(kl3mf).
```

Table 2.

Fuzzy numbers for linguistic values in determination of demand for the next year

Proposal of the 1st year	Fuzzy number	Proposal of the 1st year	Fuzzy number	Predicted demand
66	(59,66,72)	70	(63,70,77)	75
103	(93,103,113)	93	(84,93,102)	65
81	(73,81,89)	86	(78,86,94)	95

On the basis of these data, the capabilities of Data Mining technology were used to study the prediction of consumer demand. Using the FCM clustering method, clusters were obtained from the aforementioned raw data, and a fuzzy model was constructed. The clustering method was used for obtaining knowledge from the data. The fuzzy C-means method was chosen as the clustering method. In this method the sequence of calculations can be expressed as follows: The Fuzzy C-Means algorithm tries to minimize the sum of quadratic errors. The algorithm is based on iterative minimization of the following target function:

$$J(W,C)=\sum_{j=1}^{k}\sum_{i=1}^{n} w_{i.j}^{p}\, dist(x_i,c_j)^2 \qquad (10)$$

The following condition is performed for the common degree of belonging of this element x_i to all clusters:

$$\sum_{j=1}^{k} w_{i,j}=1 \qquad (11)$$

In each cluster is performed the following condition on the sum of the affiliation degrees of all elements:

$$0<\sum_{i=1}^{n} w_{i,j}<n \qquad (12)$$

The appropriate center Cj for a cluster c_j is defined as follows:

$$c_j=\frac{\sum_{i=1}^{n} w_{i,j}^{p} x_i}{\sum_{i=1}^{n} w_{i,j}^{p}}. \qquad (13)$$

The formula of updating the fuzzy separation can be obtained through minimization of the target function with a limit of weights equal 1.

$$w_{i,j} = \frac{(1/dist(x_i, c_j)^2)^{\frac{1}{p-1}}}{\sum_{q=1}^{k}(1/dist(x_i, c_q)^2)^{\frac{1}{p-1}}}. \qquad (14)$$

Clusters were obtained by the above-mentioned method and was developed an expert system. The inclusion of points in these clusters looks as follows:

a =

0.7701	0.0915	0.1385
0.4246	0.1192	0.4562
0.2250	0.1828	0.5922
0.0353	0.9145	0.0502
0.2644	0.2167	0.5188
0.2796	0.3273	0.3930
0.1773	0.5273	0.2954
0.3938	0.2950	0.3112
0.3195	0.1567	0.5238
0.2548	0.1662	0.5791
0.2298	0.5131	0.2571
0.2705	0.4564	0.2730
0.7620	0.0801	0.1580
0.9277	0.0263	0.0459
0.8215	0.0513	0.1272
0.7691	0.0498	0.1811
0.1792	0.0561	0.7647
0.0284	0.0180	0.9536
0.0576	0.0723	0.8701
0.1577	0.7158	0.1265
0.4582	0.2956	0.2462
0.7601	0.0973	0.1427

According to the results of clustering, a fuzzy model was constructed consisting of the following set of rules.

If the proposals of the previous two years 66 and 70
Then the predicted demand is 75

If the proposals of the previous two years 103 and 93
Then the predicted demand is 65
If the proposals of the previous two years 81 and 86
Then the predicted demand is 95

The reason for selecting this method is that it has several advantages over other methods. For example, this method is more similar to the k-means method. The main difference is that in the Fuzzy-Means cluster, each point has a certain degree of belonging to a certain cluster, therefore a point is not included in the "cluster" as long as it has weak Fuzzy-C Means works slower than the K-means method, due to the fact that in this method there were more calculations. Each point is evaluated for each cluster, and more operations are performed with each evaluation. K-Means is based solely on distance calculations, and this method does not calculate the degree of affiliation, i.e., it is not possible to work with inaccurate data or there is a strong cluster connection.

The fuzzy c-means method is more appropriate since it uses precise and imprecise data in predicting consumer demand.

CONCLUSIONS AND SUGGESTIONS

The main scientific results obtained in the dissertation are as follows:

1. Determined the Lyapunov index on the basis of time data of car sales for 2012 and 2013, with subsequent proof of the chaotic character of the system.
2. Calculated the fractal dimension and attractor of the dynamic system in the chaotic state.
3. Determined connection between the fractal dimension of financial time series and the application of fuzzy logic to identify the consumer demand for car sales.
4. Created fuzzy models of IF-THEN type of developed for predicting demand.
5. Performed a comparative analysis of the obtained practical results of forecasting using neural networks and c-mean page.

6. On the basis of available data, it is proved that the proposed algorithm is applicable for analysis and prediction of chaotic time series, with an uncertainty of 4.33.
7. The scientific results obtained in the dissertation might be applied to marketing in various spheres. The main practical significance lies in the possibility of determining the demand, which in turn allows budget planning and cost reduction.

List of articles published on the topic of the dissertation

1. Мустафаева С.Р. Нелинейный анализ технологической информации на объектах нефтяной и газовой промышленности / Мустафаева С.Р, Владимирский Э.И. //Оборудование и технологии для нефтегазового комплекса, Москва: - 2012. №1, стр. 48-52.

2. Мустафаева С.Р. Управление хаотическими процессами по данным наблюдений// - Баку: - Вестник Азербайджанской инженерной академии, - 2012. №4, (4) стр.57-63.

3. Мустафаева С.Р. Графические методы идентификации и прогнозирования взаимодействующих временных последовательностей // Международная научно-практическая конференция молодых ученых и студентов «Информационные процессы и технологии, - Севастополь: Изд-во СевНТУ,- 23-27 апреля, - 2012, стр.267.

4. Мустафаева С.Р. Анализ взаимодействующих хаотических временных рядов //Третья Международная научно-практическая конференция «Информационные технологии и компьютерная инженерия», - Винница: Изд-во ВНТУ 29-31 мая 2012 стр.262-263.

5. Мустафаева С.Р. Anomalous diffusion influence of chaotic system by observation // IV International conference “Problems of cybernetics and informatics”, Baku,- September 12-14, -2012, pp.158-160.

6. Мустафаева С.Р. Нелинейные методы идентификации и прогнозирования временных рядов / Мустафаева С.Р., Мамедов

Р.К., Владимирский Э.И. //Автоматизация, телемеханизация и связь в нефтяной промышленности, Москва: - 2012. №2 , - стр.17-21.

7. Мустафаева С.Р. Идентификация и прогнозирование временных рядов // Doktorantların və gənc tədqiqatçıların Azərbaycan xalqının Ümummilli lideri Heydər Əliyevin anadan olmasının 90 iliyinə həsr olunmuş” Azərbaycan 2020: neft-qaz sənayesinin inkişaf perspektivləri” konfrans ,- Bakı: - 2-3 may ,- 2013, s.183-185.

8. Мустафаева С.Р. Нелинейные методы обработки временных рядов //Международная научно-практическая конференция молодых ученых, аспирантов и студентов, - Харьков: Изд-во, 2013, стр.82.

9. Мустафаева С.Р. Анализ взаимодействия аномальной диффузии с хаотической системой по наблюдаемому временному ряду / Мустафаева С.Р., Мамедов Р.К., Владимирский Э.И. // Изв. ВТУЗ Азербайджана , - Баку: - 2013. Т. 15, №3, стр.82-88.

10. Мустафаева С.Р. Аномальная диффузия и ее влияние на хаотические системы / Мустафаева С.Р., Мамедов Р.К., Владимирский Э.И. //Информационные технологии, - Москва: - 2013. №3 ,- стр.15-19.

11. Мустафаева С.Р. Нелинейный рекуррентный анализ временных замеров объектов нефтегазодобычи // - Москва : Автоматизация, телемеханизация и связь в нефтяной промышленности , - 2016. . №8 , стр.40-42.

12. Мустафаева С.Р. Кросс-рекуррентный анализ коррелированных систем в нефтегазодобыче // - Москва: Автоматизация, телемеханизация и связь в нефтяной промышленности - 2016, №2 стр.3-25.

13. Мустафаева С.Р. Аддитивное и мультипликативное взаимодействие стохастических процессов // - Баку: Вестник Азербайджанской Инженерной Академии - 2017, № 2(9), стр.115-120.

14. Мустафаева С.Р. Использование метода кластеризации нечеткой логики в вопросах прогнозирования в нефтяной

индустрии / Мустафаева С.Р., Н.Г. Джавадов // Автоматизация, телемеханизация и связь в нефтяной промышленности, Москва : - 2019. № 7(552) стр.49-52.

15. Mustafayeva S.R. Calculation of Lyapunov exponent of time series //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-2019 Switzerland, pp.768-771.

16. Mustafayeva S.R. Fractal Dimension Determining for Demand Forecasting //10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions – Prague: - August 27–28, 2019 ICSCCW-Switzerland, pp.931-937.

17. Мустафаева С.Р.Использование метода кластеризации в автомобильном бизнесе, ADNSU-nun 100 illiyinə həsr olunmuş ADNSU gənc alim və tədqiqatçıların elmi-praktiki konfransı 2020-Bakı: - 2020 ADNSU, стр.452-457.

18. Mustafayeva S.R. Using the clustering method in the automotive business // Bakı: Azərbaycan Ali Texniki Məktəblərinin Xəbərləri, Bakı: - 2022. s.33-37.

Personal contribution of the applicant in the works published in co-authorship:

[2,3,4,5,7,8,11,12,13,15,16,17,18] - self-conducted studies;

[1,6,9] - nonlinear analysis of information flows in information and measuring systems was carried out;

[4,8,12] - nonparametric analysis of chaotic time series;

[15,16,17,18] - proposed research in the automotive field using fuzzy logic methods.

The defense will be held on 16.09.2022 at 13.00 at the meeting of Dissertation council_ED2.02_of Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at Azerbaijan State Oil and İndustry University

Address: AZ1010, Azerbaijan State Oil and İndustry University,main building, room # 250

Dissertation is assemble at the Azerbaijan State Oil and İndustry University Library

Electronic versions of dissertation and its abstract are available on the official website o the Azerbaijan State Oil and İndustry University

Abstract was sent the reqired addresses on 22.07.2022

Signed for print: 30.06.2022
Paper format: A5
Volume: 40285
Number of hard copies: 20