A Method of Generating Fuzzy Sets from Homogeneous and Monotonous Time Series

Dmytro V. Stefanyshyn[1], Yulija D. Stefanyshyna[1]

[1] *The Institute of Telecommunications and Global Information Space of the National Academy of Sciences of Ukraine, 13, Chokolivsky av., Kyiv, 03186, Ukraine*

dvstefanyshyn@yahoo.com, stefanishina@yahoo.com

Abstract. *The method of overcoming the uncertainty of the extrapolation prediction based on time series analysis is proposed. Homogeneous and monotonous time series are analyzed. The research effort is to establish a weighted average expected value of the forecasted parameter. Forecasting is carried out within the framework of fuzzy modeling. To solve the problem a number of model trends based on different reference point samples of the observed time series is built. These trends are regarded as expert estimates to generate the fuzzy sets of values of the fuzzy variables. In the study deals with such fuzzy variables as "expected value of the variable will not be less than…", "expected value of the variable will be no more than …", "expected value of the variable will be in the interval…". Membership functions of fuzzy variables to proper fuzzy sets are built by values of determination coefficients of selected trends. The problem is solved by an example of forecasting of expected annual damage due to floods in the Tisza River Basin in the Transcarpathian region of Ukraine for the period until 2020.*

Keywords

Fuzzy modeling, fuzzy variable, homogeneous and monotonous time series, membership function, trend.

1 Introduction

In general, time series have several nonrandom components depending on time. Nonrandom components of homogeneous and monotonic time series are usually described by means of trends.

The standard approach to the analysis of time series implies a search for the best trend satisfies the requirements of balancing available data [1, 2]. But even the best trend is merely a pattern of an expected value behavior depending on time within the time series of observed data. Therefore, the forecasting problem of expected values of random variable depending on time, based on trends extrapolations, should be viewed as a task burdened by uncertainty.

2 Presentation of the method

2.1 Some of the terms of the fuzzy sets theory

Methods of the fuzzy sets theory presently are widely used for data formalization, which presented in the form of expert evaluations (judgments) [3, 4] with introducing the conception of linguistic variable. Exemplification of the linguistic variable can be such expressions as "approximately equal to …", "is in the interval …", "is similar to…" etc.

Formally fuzzy set $\tilde{A}$ defines as a set of ordered pairs (tuples) presented as $< x, \mu_A(x) >$, where x is the element of some universal set X, and $\mu_A(x)$ is a membership function of x to $\tilde{A}$. The membership function $\mu_A(x)$ assigns each element $x \in X$ some real number in the interval of [0, 1]: $\mu_A : X \to [0,1]$. If $\mu_A(x) = 1$, then x definitely belongs to $\tilde{A}$, if $\mu_A(x) = 0$, then x definitely does not belong $\tilde{A}$.

The membership function $\mu_A(x)$ can be the arbitrary function, which is given analytically in the form of function $f(x)$ or graphically in the form of some curve.

There are wide varieties of practical representations of membership functions: triangular and trapezoidal membership functions; Z -similar and S -similar spline functions; sigmoid, bell-curve functions; linear Z -similar and S -similar functions. It can be also classical probability distribution functions, e.g. the function of normal distribution law etc. Selection of membership functions can be determined by their practical utilizing. This procedure is called fuzzification or brining to the illegibility. The purpose of this procedure is to generate fuzzy sets when a variable is modeled as a fuzzy number or a fuzzy interval. Experts can directly determine membership functions for fuzzy parameters, or the fuzzification procedure is realized by means of fuzzy linguistic variables.

Reduction of a fuzzy parameter x to its crisp value $\hat{x}$ is carried out using so-called defuzzification procedure. Commonly used the weighting method (centroid method):

$$\hat{x} = \frac{\int_{Min}^{Max} x \cdot \mu_A(x) dx}{\int_{Min}^{Max} \mu_A(x) dx}, \quad (1)$$

where *min*, *max* are the left and the right values of the carrier interval of a fuzzy set $\tilde{A}$ which determine range of definition of x; $\mu_A(x)$ is the membership function of x to the fuzzy set $\tilde{A}$.

2.2 General statement of the problem

Let there be monotonous and homogeneous time series. Let time series component, which depends on the time, modeled by various trends. Trends can be selected appropriately, for example, based on different sampling intervals of the time series. As a result of forecasting using trends some expected values of the parameter can be obtained. Weight of each of these values may be determined by the determination coefficient of the particular trend. These values can later be regarded as expert estimates of the parameter, which is predicted.

Let forecasting carried out within the framework of fuzzy modeling. To solve the problem the model trends based on different reference point samples of the observed time series are regarded as expert estimates to generate fuzzy sets of values of fuzzy variables.

2.3 An example of generation of fuzzy sets based on the trends

As an example the forecasting problem of expected annual damage induced by floods in the Tisza River Basin in the Transcarpathian region of Ukraine was considered. The annual damage induced by floods in the region since 1955 is given in table 1.

The perspective forecasting of expected damage y carried out for the period until 2020. For the forecasting purpose, trends were built for different reference point sample series. MS Excel was used for trends modeling. Selection of trends was realized in the countdown with the retrospective "deepening".

It was found that determination coefficients of trends, as far as additional retrospective data taking into account, increases up regular to the trend with the reference point at 1970, when there was decreasing of determination coefficient occurred in comparison with 1971 trend. Afterwards, recurring increases of determination coefficients with the reference point from 1958 to 1970 were observed. With consideration for such determination coefficients "behavior" two groups of trends were selected:

A) trends with the reference point on the interval from 1971 to 1986 (see fig. 1);

B) trends with the reference point on the interval from 1958 to 1970 (see fig. 2).

It was accepted that each of the modeling trends represents a certain expert judgment of the expected damage for perspective, and measures of confidence to the expected damage y estimates on different trends were found by their determination coefficients.

Tab. 1. Annual damage y from catastrophic floods in the Tisza River Basin in the Transcarpathian region of Ukraine (according to the State committee of water management in Ukraine)

Year	Damage , mil. hr.	Year	Damage, mil. hr.
1955	43,7	1974	33,6
1957	21,3	1975	87,9
1958	9,8	1976	60,1
1965	14,2	1977	85,2
1967	10,9	1978	115,8
1968	10,4	1979	64,4
1969	34,4	1980	325
1970	237	1986	127,9
1971	35,5	1998	81
1972	24	2001	284,9
1973	23,5	2008	163

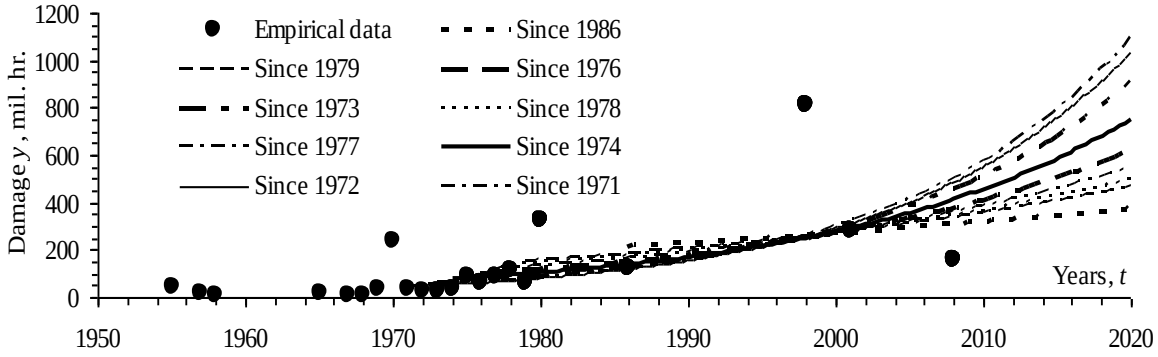

Fig. 1. Trends with the reference point on the interval from 1971 to 1986 (data from table 1)

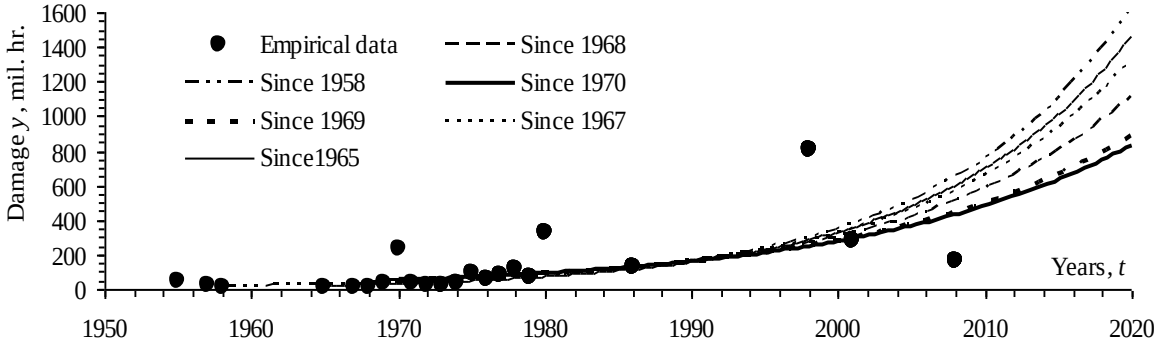

Fig. 2. Trends with the reference point on the interval from 1958 to 1970 (data from table 1)

Further, expected damage y estimates, which were obtained by trends with the reference point on the interval from 1971 to 1986 and by trends with the reference point on the interval from 1958 to 1970 were related to different fuzzy sets $\tilde{A}$, $\tilde{B}$, and linguistic variables were introduced.

For the fuzzy set $\tilde{A}$ built according to forecasting results with the reference point on the interval from 1971 to 1986 the linguistic variable was determined as "*expected annual damage to 2020 will be not less than…*". For the fuzzy set $\tilde{B}$ built according to forecasting results with the reference point on the interval from 1958 to 1970 the linguistic variable was determined as "*expected annual damage to 2020 will be not more than…*". Values of membership functions $\mu_A(y_i)$, $\mu_B(y_i)$ of expected annual damage y_i formed fuzzy sets $\tilde{A}$, $\tilde{B}$ were determined according to determination coefficients of appropriate trends, namely:

$$\mu_A(y_i) = R_i^2 \big/ R_{i,\max}^2 , \tag{2}$$

$$\mu_B(y_i) = 1 - R_i^2 \big/ R_{i,\max}^2 , \tag{3}$$

where R_i^2 is a determination coefficient of an i-th trend; $R_{i,\max}^2$ is the maximum value of determination coefficients among distinguished groups of trends.

Results of expected damage forecasting are given in tables 2, 3. Membership functions of the corresponding linguistic variables are shown below in fig. 3, 4.

Tab. 2. Expected annual damage after forecasting with using data from 1971 to 1986

Trend	Trend equation for expected damage y	Expected damage y, millions, hr.	R_i^2	$\mu_A(y)$
1986	$y = 3 \cdot 10^{-106} \cdot t^{32,703}$	373,3	0,0336	0,061674
1979	$y = 6 \cdot 10^{-180} \cdot t^{55,029}$	465,9	0,1446	0,265419
1978	$y = 9 \cdot 10^{-201} \cdot t^{61,338}$	498,8	0,2038	0,374082
1977	$y = 2 \cdot 10^{-232} \cdot t^{70,936}$	588,1	0,2768	0,508076
1976	$y = 4 \cdot 10^{-272} \cdot t^{82,955}$	627,3	0,3511	0,644457
1974	$y = 2 \cdot 10^{-41} \cdot \exp(0,04967 \cdot t)$	750,4	0,4352	0,798825
1973	$y = 3 \cdot 10^{-48} \cdot \exp(0,05754 \cdot t)$	902,6	0,4609	0,845999
1972	$y = 5 \cdot 10^{-53} \cdot \exp(0,06306 \cdot t)$	1046,9	0,5154	0,946035
1971	$y = 1 \cdot 10^{-54} \cdot \exp(0,06502 \cdot t)$	1097,5	0,5448	1

Tab. 3. Expected annual damage after forecasting with using data from 1958 to 1970

Trend	Trend equation for expected damage y	Expected damage y, millions, hr.	R_i^2	$\mu_B(y)$
1970	$y = 5 \cdot 10^{-46} \cdot \exp(0,05497 \cdot t)$	837,1	0,3974	0,31317
1969	$y = 7 \cdot 10^{-48} \cdot \exp(0,05711 \cdot t)$	883,6	0,4289	0,258728
1968	$y = 2 \cdot 10^{-54} \cdot \exp(0,06468 \cdot t)$	1104,5	0,4609	0,203422
1967	$y = 1 \cdot 10^{-59} \cdot \exp(0,07081 \cdot t)$	1317,7	0,4945	0,145351
1965	$y = 1 \cdot 10^{-62} \cdot \exp(0,07428 \cdot t)$	1458,7	0,5288	0,08607
1958	$y = 7 \cdot 10^{-65} \cdot \exp(0,07678 \cdot t)$	1593,2	0,5786	0

It was found that the membership function of the expected damage y represented as the fuzzy variable "*expected annual damage to 2020 will be not less than…*" within values from 0 to 1 is well approximated by the polynomial:

$$\mu_A(y) = 2 \cdot 10^{-9} \cdot y^3 - 6 \cdot 10^{-6} \cdot y^2 + 0,006705 \cdot y - 1,7537 , \tag{4}$$

with the determination coefficient $R^2 = 0{,}9916$, and values of $y \le 360$ millions hr. when $\mu_A(y) = 0$.

The membership function of the expected annual damage y represented as the fuzzy variable "*expected annual damage to 2020 will be not more than…*" within the function values from 0 to 1 is well approximated by the function:

$$\mu_B(y) = -0{,}0004 \cdot y + 0{,}65, \tag{5}$$

with the determination coefficient $R^2 = 0{,}9706$, and values of $y \ge 1625$ millions hr. when $\mu_B(y) = 0$

Further a linguistic variable "*expected annual damage to 2020 will be in the interval…*" was considered and a fuzzy set $\tilde{C}$ was found: $\tilde{C} = \tilde{A} \cap \tilde{B}$. The membership function of the linguistic variable "*expected annual damage to 2020 will be in the interval…*" is shown in fig. 5. The point of intersection (y = 536, $\mu_C(y)$ = 0,4356) of the membership functions $\mu_A(y)$, $\mu_B(y)$ was found by solving the system of equations:

$$\begin{cases} \mu_A(y) = 2 \cdot 10^{-9} \cdot y^3 - 6 \cdot 10^{-6} \cdot y^2 + 0{,}006705 \cdot y - 1{,}7537 \\ \mu_B(y) = -0{,}0004 \cdot y + 0{,}65 \end{cases}. \tag{6}$$

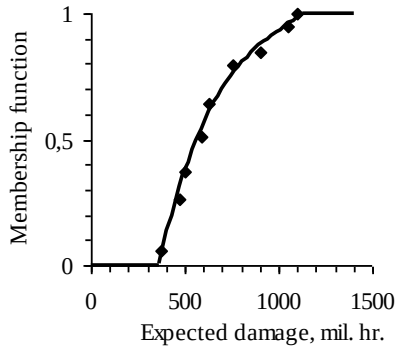

Fig. 3. The membership function of the fuzzy set $\tilde{A}$

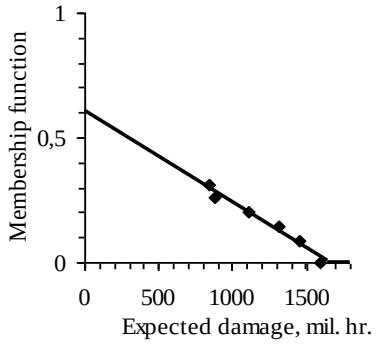

Fig. 4. The membership function of the fuzzy set $\tilde{B}$

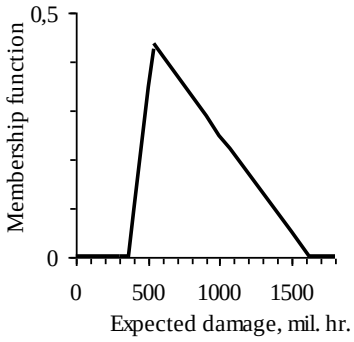

Fig. 5. The membership function of the fuzzy set $\tilde{C} = \tilde{A} \cap \tilde{B}$

In the end, after the defuzzification procedure according to (1) the crisp value of the expected annual damage $\hat{y}$ = 980,15 millions hr. on a time horizon to 2020 was obtained.

3 Conclusion

The method of the fuzzy sets generating from homogeneous and monotonous time series in forecasting problem based on extrapolation trends was presented. The method allows to overcome the uncertainty of the extrapolation prediction based on time series analysis. Forecasting is carried out within the framework of fuzzy modeling. To solve the problem a number of model trends based on different sampling intervals of the observed time series is proposed to build. All of these trends can be regarded as expert estimates to generate fuzzy sets of values of fuzzy variables. Measures of confidence of these expert estimates whereby the membership functions were built were found by trends determination coefficients. The practical problem for forecasting example of expected annual damage from floods in the Tisza River Basin in the Transcarpathian region of Ukraine to 2020 was reviewed.

References

[1] Brockwell P.J., Davis R.A. Time series: Theory and methods. Second edition. Springer, 1991. 567 p.
[2] Hamilton J.D. Time series analysis. Princeton University Press. Princeton, New Jersey, 1994. 782 p.
[3] Klir G.J., Yuan Bo. Fuzzy Sets and Fuzzy Logic: Theory and Applications. Prentice Hall. PTR, 1995. 592 p.
[4] Zimmermann H.-J. Fuzzy set Theory and its applications. Springer. 4-th edition, 2001. 544 p.