Global and Stochastic Analysis
Vol. 8 No. 1 (January-June, 2021)

MUK PUBLICATIONS
Open Access Publisher

THE PROBABILITIES OF FUZZY EVENTS IN THE STUDY OF THE STRUCTURE OF SEDIMENTARY ROCKS

V.YU. ITKIN AND V.M. SINITSYNA

Abstract. A problem of automatic classification of rock fragments is investigated. A rock fragment image is characterized by two numeric indicators, elongation and form-factor, which make it possible to assign this fragment to one of the 9 categories (groups). The debris formation in the process of destruction of rocks and accumulation of sediment is of a random nature, therefore the problem of calculating of probabilities of getting a sample into a certain group arises. However, there are no crisp boundaries between the categories of fragments, allowing to unambiguously assign the sample to one or another group. Therefore, to solve classification problems and calculate probabilities the methods of fuzzy logic and mathematical statistics were jointly applied.

The probability of a fuzzy event can be characterized by either a single number or a function. The functional probabilistic characteristic is defined as the dependence of the crisp probabilities of α-cuts on the degree of membership α. To calculate the numerical probabilistic characteristics, two approaches are considered, similar to two defuzzification methods: centroid and bisector. The first method is correct if the linguistic variable is consistent, i.e. it satisfies the normalization condition. The second one is correct for variables that admit a partition of the universe by α-cuts into nonintersecting domains.

The application of the described methods to the problems of classification of fragments made it possible to obtain different probabilistic characteristics of all fuzzy groups. The probabilities obtained by different methods turned out to be similar. As a result of the analysis, the most probably categories for the considered samples were identified. They turned out to be groups of semi-rounded isometric and elongated rock fragments.

The proposed approach makes it possible to compare rock samples for different areas of a field and identify areas with the greatest similarity or difference in rock structure.

Introduction

Uncertainty of statistical information can be caused by various reasons. If knowledge about a phenomenon is not enough to describe it reliably, models of

Date: Date of Submission October 5, 2020; Date of Acceptance November 24, 2020, Communicated by Yuri E. Gliklikh.

2010 *Mathematics Subject Classification.* Primary 60A86; Secondary 62A86, 62E86.

Key words and phrases. Fuzzy event; probability; alpha-cut; two-dimensional membership function; classification of the shape of fragments; roundness; elongation.

* The authors are grateful to I.I. Khasanov, who had formulated the problem of classification of rock fragments and provided the thin section image, and Prof. V.V. Rykov, who carefully read the article and made valuable comments.

random events and variables are used. Statistical information is a data for estimating of probabilities of random events and distribution laws of random variables. Another reason for uncertainty is the fuzziness of intuitive qualitative ideas about the subject of research. In lithology, when studying the structure of sedimentary rocks, the concepts of *roundness* and *elongation* of fragments are used. When classifying, the shape of a fragment is compared with the intuitive standard, an example of which is shown in Fig. 1.

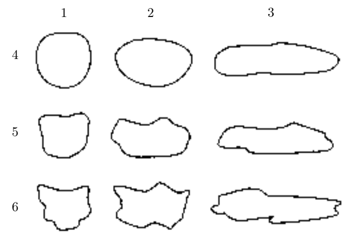

FIGURE 1. The shapes of rock fragments [5] (1 – isometric, 2 – elongated, 3 – sharply elongated, 4 – rounded, 5 – semi-rounded, 6 – non-rounded).

However, it is impossible to draw crisp boundaries between the categories of rock fragments, therefore, for classification, it is natural to use the apparatus of fuzzy sets [9], which is successfully applied in various geological studies [2, 10, 12].

Fuzziness and randomness are different types of uncertainty. Although the properties of fuzzy sets can be obtained from the properties of random sets [7], in applied research these types of uncertainty have different interpretations [1] and can be used together [14]. A detail review of various methods of uncertainty modelling in geology is given in [4].

Fuzzy events and their probabilities are first considered by the founder of fuzzy logic L. Zadeh [13]. Various mathematical aspects of the theory of fuzzy random events have been studied by many authors [3, 6, 11]. The probabilities of fuzzy events are also calculated in engineering papers [8].

In this paper, a new approach to the calculating of fuzzy probabilities is proposed, illustrated with examples related to the classification of rock fragments.

1. Fuzzy event probability

According to the generally accepted approach, a fuzzy event A is characterized by one indicator – a probability, which is calculated as the mathematical expectation of the membership function $\mu(x)$ of the corresponding fuzzy set A, all elements of which are assigned some probabilities. Let a random variable X, determined by the cumulative distribution function $F(x)$, can take all possible values of the universe Ω. Then the probability (or *mean probability*) of a fuzzy event is equal to

$$\overline{P} = \mathbf{P}(A) = \mathbf{E}(\mu(X)) = \int\limits_{\Omega} \mu(x)\, dF(x). \tag{1.1}$$

This calculation method is similar to the centroid method of defuzzification of a fuzzy set.

Averaging leads (except for trivial cases) to the loss of information, therefore, along with probability (1.1) it is convenient also to have a functional characteristic of a fuzzy event, showing probabilities for all degrees of membership of a fuzzy set – an analogue of the cumulative distribution function $F(x)$ or the survival function $R(x) = 1 - F(x)$ for classical (crisp) random variables.

Let us set some degree of membership $\alpha \in [0, 1]$ and construct the α-cut

$$A_\alpha = \{x \in A : \mu(x) \geq \alpha\}.$$

Its crisp probability is equal to

$$P(\alpha) = \mathbf{P}(A_\alpha) = \int\limits_{A_\alpha} dF(x) = \int\limits_{\Omega} 1\{\mu_A(x) \geq \alpha\}\, dF(x). \tag{1.2}$$

The expression (1.2) gives a functional probabilistic characteristic of a fuzzy set.

Let a collection of fuzzy sets $\{A^k\}$ with membership functions $\mu_k(x)$ determine a linguistic variable a in the universe Ω, i.e. $\{A^k\}$ is a term collection. Then it is naturally to require the normalization condition for the mean probabilities (1.1),

$$\sum_k \overline{P}_k = 1, \tag{1.3}$$

where $\overline{P}_k = \mathbf{P}(A^k)$ are probabilities of the terms A^k.

This condition is true if the variable a is *consistent*, i.e. for all fixed $x \in \Omega$ the equality

$$\sum_k \mu_k(x) = 1 \tag{1.4}$$

is satisfied.

The definition (1.2) allows us to introduce another numerical characteristic of the probability of a fuzzy event, similar to the bisector defuzzification method of a fuzzy set. Let us denote the term probabilities calculated by formula (1.2) as $P_k(\alpha)$. For a fixed α, the total probability $\sum\limits_k P_k(\alpha)$ can be either greater or less than 1. However, if $\alpha = q$ exists, for which a collection of α-cuts $\{A^k_q\}$ is a partition of universe Ω into nonintersecting domains, i.e.

$$\begin{aligned} \mathop{\cup}_k A^k_q &= \Omega, && (1.5) \\ A^i_q \cap A^j_q &= \varnothing, \quad i \neq j, && (1.6) \end{aligned}$$

then the normalization condition is satisfied for $P_k(q)$,

$$\sum_k P_k(q) = 1. \tag{1.7}$$

For symmetry, it is recommended to construct membership functions so that $q = 0.5$.

Thus, it is possible to calculate the mean probabilities $\overline{P}_k$ for the consistent linguistic variables. For variables, allowing some partition with α-cuts, it is possible to calculate the probabilities $P_k(q)$. The linguistic variables with both qualities have although similar, but different numerical characteristics of probabilities.

2. Classification of rock fragments in a thin section of sandstone

As an example we present the results of processing statistical information on the shape of rock fragments and grains in a thin section of sandstone, some of them have already been described in [9].

Plane images of rock fragments have random characteristics, for instance, form-factor f (ratio of area to perimeter squared) and elongation l (ratio of length to width). These characteristics make it possible to classify the fragments by shape and assign them to one of the groups schematically shown in Fig. 1.

Elongation and form-factor, as established in [9], are distributed according to the lognormal law. These random variables can be considered independent, because the correlation of their logarithms is insignificant according to the Students test,

$$\begin{aligned} \hat{\rho}(\ln f, \ln l) &= 0.08, \\ p &= 0.16. \end{aligned}$$

However, variables f and l can be considered also as linguistic variables with one-dimensional membership functions $\mu_{if}(f)$ for form-factor and $\mu_{jl}(l)$ for elongation, $i, j = \overline{1, 3}$ (for details see [9]).

3. One-dimensional problem of calculating probabilities

On the one hand, elongation l and form-factor f are linguistic variables that are described by the terms membership functions *low*, *average* and *high*. They are shown in Fig. 2 in solid lines; their ordinate axes are on the left. However, on the other hand, l and f are random variables that are described by lognormal probability density functions. They are shown in Fig. 2 in dash lines, their ordinate axes are on the right.

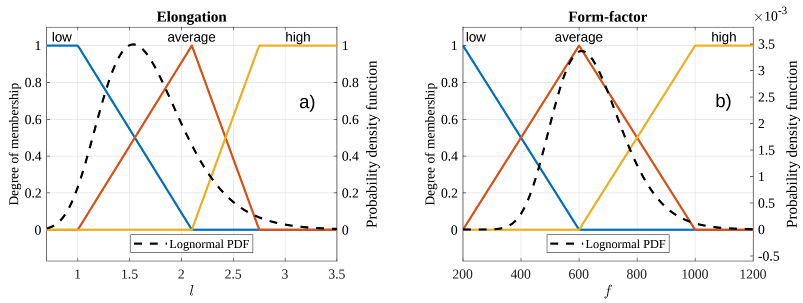

FIGURE 2. Membership functions and probability density functions of elongation and form-factor

A collection of membership functions is a fuzzy description of a variable formed by an expert. The statistical data influences expert opinion, but does not completely determine it. It depends on personal experience, the meanings of the concepts "low", "average" and "high" in natural language, the ability to visualize, and other subjective factors.

Fig. 2a shows that the mode of the probability density function of the elongation is located significantly to the left of the core of the average term. Moreover, Fig. 2b shows that these indicators for the form-factor are almost equal. That is, the statistical data is close to intuitive ideas about the average form-factor and differ significantly from them for the average elongation.

For the considered fuzzy sets, α-cuts are intervals of the form $\big[a(\alpha), b(\alpha)\big]$, therefore, the probabilities of getting into them can be found analytically,

$$P(\alpha) = F\big(b(\alpha)\big) - F\big(a(\alpha)\big), \tag{3.1}$$

where $F(x)$ is a cumulative distribution function of a random variable.

The same probabilities can be estimated from empirical data,

$$\hat{P}(\alpha) = \frac{N\{\mu(x_i) \geq \alpha\}}{n}, \tag{3.2}$$

where n is the sample size, $N\{\mu(x_i) \geq \alpha\}$ is the count of observations x_i in the α-cut.

The figures 3 and 4 show the plots of the probabilities of the terms for the elongation and the form-factor.

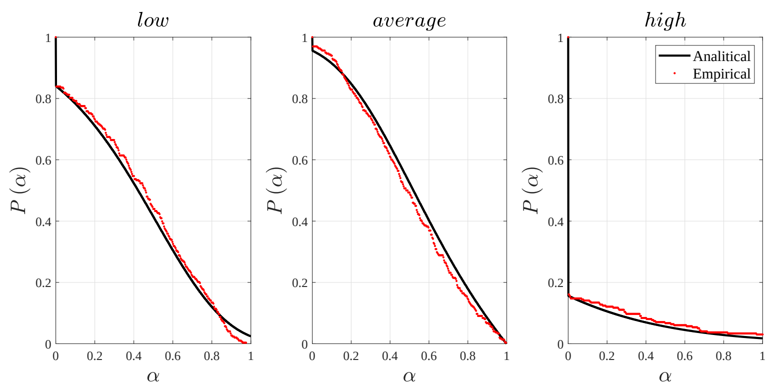

FIGURE 3. Probabilities of the elongation's terms

The considered linguistic variables are consistent, and if $\alpha = 0.5$ then their terms α-cuts partition the universe into nonintersecting domains. Therefore, both indicators, $\overline{P}$ and $P(0.5)$, are correct numerical probabilistic characteristics of fuzzy sets. Table 1 shows these indicators, calculated both by analytical formulas $\big(\overline{P},\ P(0.5)\big)$ and by empirical data $\big(\hat{P},\ \hat{P}(0.5)\big)$.

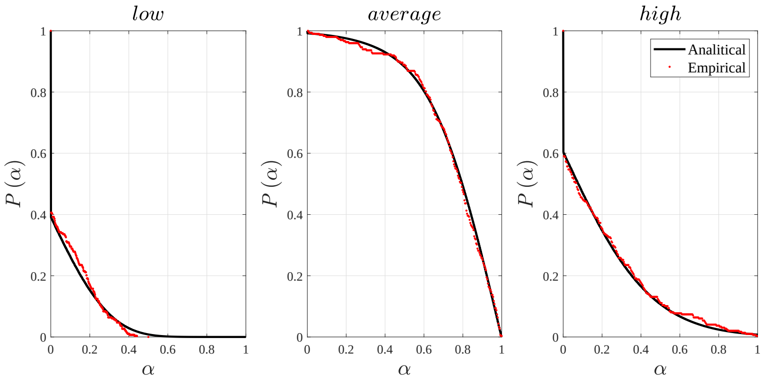

FIGURE 4. Probabilities of the form-factor's terms

TABLE 1. Probabilities of the terms for the elongation and the form-factor.

Variable	Elongation l			Form-factor f		
Term	*low*	*average*	*high*	*low*	*average*	*high*
$\overline{P}$	0.419	0.514	0.067	0.072	0.749	0.179
$\hat{P}$	0.432	0.490	0.078	0.074	0.743	0.183
$P(0.5)$	0.417	0.525	0.058	0.008	0.883	0.109
$\hat{P}(0.5)$	0.440	0.493	0.067	0	0.889	0.111

Table 1 shows that all the methods give close results, moreover:

- the probabilities of average and low elongation are high enough;
- the average is the most probably value of form-factor;
- the probabilities of the low elongation and the low form-factor are negligible;
- the probability of the high form-factor is nonzero but not too large.

4. Two-dimensional problem of calculating of probabilities

Solving the problem of the rock fragments classification, we use both the elongation and form-factor, and therefore, it is necessary to consider two-dimensional fuzzy sets. Depending on values of the elongation and the form-factor, a rock fragment can be assigned with different degrees of membership to each of the 9 groups (Fig. 1), which are denoted by $G^k, k = \overline{1,9}$.

Lets calculate the probabilities of getting a rock fragment into each group. For this, we will:

- consider each group G^k as a two-dimensional fuzzy set and construct the membership function $\mu_k(f,l)$ for it; all these functions will determine the linguistic variable $g = l \wedge f$, which is the disjunction of the elongation and the form-factor;
- find α-cuts G^k_α of the groups G^k;

- calculate the probabilities of getting of the pair (f, l) into each α-cut G_α^k for a fixed α.

Lets consider the stages of this algorithm in more details.

To assign a rock fragment into some group G^k it is necessary the form-factor and the elongation were elements of the terms determined by the membership functions μ_{if} and μ_{jl} accordingly. That's why the two-dimensional membership functions of the group G^k should be constructed with the fuzzy disjunction of the one-dimensional membership functions of variables f and l,

$$\mu_k(f,l) = \mu_{if}(f) \cdot \mu_{jl}(l) \tag{4.1}$$

or

$$\mu_k(f,l) = \min\big(\mu_{if}(f), \mu_{jl}(l)\big), \quad k = \overline{1,9}. \tag{4.2}$$

Both linguistic variables, f and l, are consistent, therefore if the disjunction is calculated through the product by the formula (4.1), then the variable g is also consistent (Fig. 6),

$$\sum_k \mu_k(f,l) = \sum_i \sum_j \Big(\mu_{if}(f) \cdot \mu_{jl}(l)\Big) = \sum_i \left(\mu_{if}(f) \sum_j \mu_{jl}(l)\right) = 1. \tag{4.3}$$

However, α-cuts collection is not a partition of the universe for any α (Fig. 7a): if $\alpha = 0.5$ (shown in Fig. 7a in red dash lines), then some α-cuts have the common borders, but there are some domains located outside any α-cuts. If $\alpha = 0.4$ then there are both α-cuts intersections (dark green areas in Fig. 7a) and domains outside any α-cuts (white areas in Fig. 7a).

If the operation of minimum by formula (4.2) is used for the disjunction calculation then the variable g is inconsistent (Fig. 5),

$$\sum_k \mu_k(f,l) = \sum_i \sum_j \min\Big(\mu_{if}(f), \mu_{jl}(l)\Big) \neq 1, \tag{4.4}$$

but if $\alpha = 0.5$ then the α-cuts of the terms G^k are a partition of the universe into nonintersecting domains (red dash lines in Fig. 7b).

Therefore, every method of disjunction calculation allows obtaining only one numerical characteristic of a fuzzy event probability.

Lets calculate the probabilities of a two-dimensional fuzzy event obtained by the minimum operation (4.2). In this case α-cuts G_α^k are rectangles (Fig. 7b),

$$G_\alpha^k = \Big[f_{\min}^k(\alpha), f_{\max}^k(\alpha)\Big] \times \Big[l_{\min}^k(\alpha), l_{\max}^k(\alpha)\Big] \tag{4.5}$$

As the variables f and l are independent, the probability of getting into an α-cut can be obtained by a product of the probabilities of getting of the values f and l into corresponding intervals. The lognormal law parameters for random variables f and l can be estimated by statistical data, thats why the cumulative distribution functions $F_f(f)$ and $F_l(l)$ are known. Therefore, the probability of

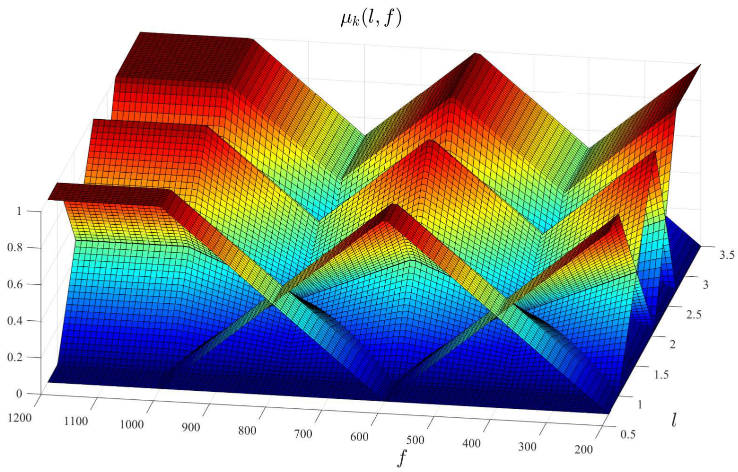

FIGURE 5. Membership functions obtained by the product operation (4.1).

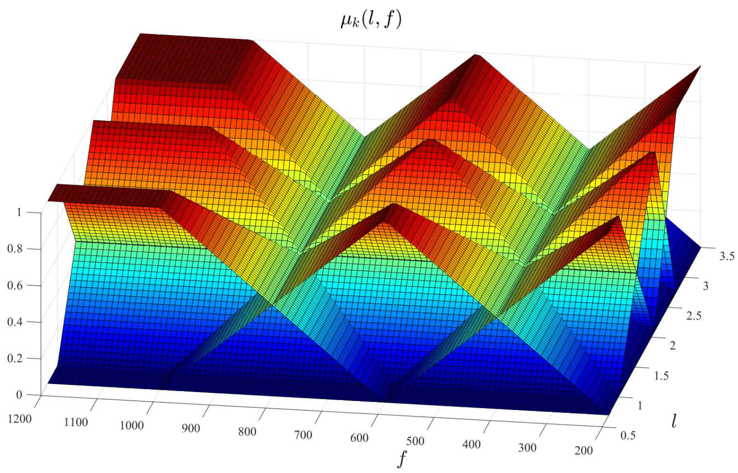

FIGURE 6. Membership functions obtained by the minimum operation (4.2).

getting into α-cut G_α^k is equal to

$$P_k(\alpha) = \left[F_f\left(f_{\max}^k(\alpha)\right) - F_f\left(f_{\min}^k(\alpha)\right)\right] \cdot \left[F_l\left(l_{\max}^k(\alpha)\right) - F_l\left(l_{\min}^k(\alpha)\right)\right]. \quad (4.6)$$

The same probability can be estimated by experimental data. Lets determine the count of the rock fragments, the membership function of which $\mu_k(f,l)$ is

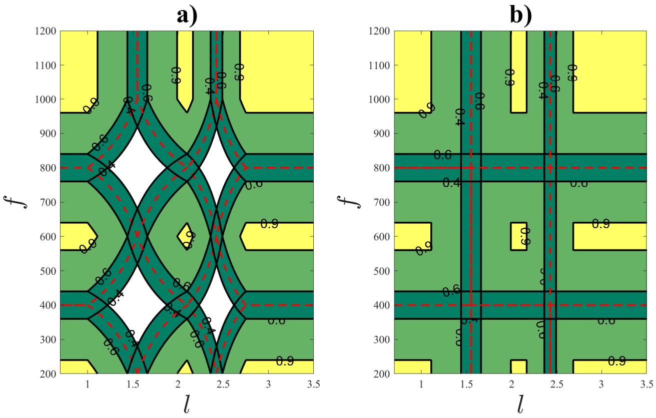

FIGURE 7. α-cuts obtained by two methods of disjunction calculation: product **(a)** and minimum **(b)**.

greater or equal α, and divide it by the total rock fragments count n,

$$\hat{P}_k(\alpha) = \frac{N\{\mu_k(f,l) \geq \alpha\}}{n}. \tag{4.7}$$

Fig. 8 shows the results of calculation of probabilities by (4.6) and (4.7).

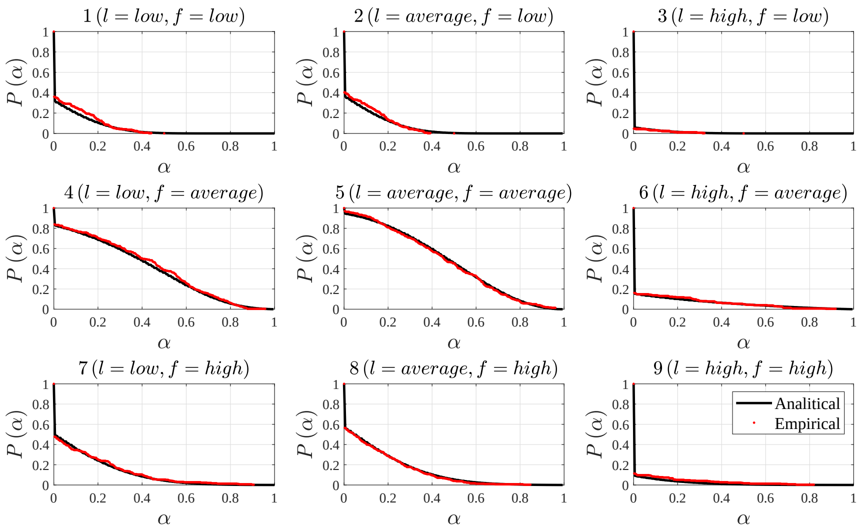

FIGURE 8. Probabilities of getting rock fragments into fuzzy groups

For $\alpha = 0.5$, α-cuts $G^k_{0.5}$ obtained by the minimum operation (red dash lines in Fig. 7b) do not intersect and their union is equal to the universe. Therefore, for

$\alpha = 0.5$ the normalization condition is satisfied, and $P_k(0.5)$ is a correct numerical probabilistic indicator of the term G^k.

In Table 2, the values of this indicator are compared with the mean probabilities calculating both by analytical formulas $(\overline{P}_k)$ and empirical data $(\hat{P}_k)$. In the first case the formula (1.1) was used and in the second one the following from it formula

$$\hat{P}_k = \frac{1}{n}\sum_i \mu_k(f_i, l_i) \tag{4.8}$$

was used.

The mean probabilities were calculated by the membership functions obtained by the product operation (4.1), as just in this case, the linguistic variable g is consistent and the normalization condition (4.3) is satisfied.

TABLE 2. Probabilities of getting into fuzzy groups

k	1	2	3	4	5	6	7	8	9
$\overline{P}_k$	0.030	0.037	0.005	0.314	0.385	0.050	0.075	0.092	0.012
$\hat{P}_k$	0.034	0.036	0.004	0.320	0.372	0.051	0.078	0.083	0.022
$P_k(0.5)$	0.003	0.004	0	0.366	0.455	0.049	0.053	0.054	0.016
$\hat{P}_k(0.5)$	0	0	0	0.386	0.453	0.050	0.054	0.040	0.017

Table 2 shows that the probabilities calculated by various methods are close enough, whereby the groups 4 (semi-roundness isometric fragments) and 5 (semi-roundness elongated fragments) are most probably.

Conclusion

A rock fragment shape is characterized by two parameters: the elongation and the form-factor. After measuring them the fragment can be assigned to one of the 9 groups. The groups have no crisp borders, thats why the fuzzy logic methods should be used for the classification, and it is necessary to combine fuzzy logic with probability theory and mathematical statistics for calculating probabilities of getting a fragment into one of the groups.

Getting a fragment into a group is a fuzzy event, which can be described both by numerical and functional probability characteristics. The functional characteristics are defined by calculating of the crisp probabilities for a sequence of the α-cuts. The numerical probabilities can be obtained by both the methods:

- calculating of the mean of a membership function;
- calculating of the crisp probabilities for α-cuts with a fixed degree of membership providing a partition of the universe into nonintersecting domains.

The first method is correct for consistent linguistic variables, and the second method is only prospered for variables allowing a partition of the universe into nonintersecting domains.

Applying of the described methods to the problems of the rock fragments classification has allowed to obtain the probabilistic characteristics of the fragment groups both for one-dimensional and two-dimensional cases. The analysis of the available thin rock section has showed:

- the average and low elongation have the probabilities close to 0.5;
- the high elongation is improbable;
- the average form-factor is most probably;
- the groups of semi-roundness isometric and elongated fragments are most probably.

The proposed method can be applied to compare rock samples at different field areas and identifying areas with the greatest similarity or difference in rock structure.

References

1. I. Aliev, M., A. Isaeva, E., and M. Aliev, I., *Teoriya veroyatnostej i teoriya nechetkih mnozhestv l. zadeh: razlichiya i skhodstvo [probability theory and theory of fuzzy sets l. zadeh: differences and similarities]*, Iskusstvennyj intellekt i prinyatie reshenij [Artificial intelligence and decision making] **3** (2012), 19–25.
2. A. Aydin, *Fuzzy set approaches to classification of rock masses*, Engineering Geology **74** (2004), 227–245.
3. J. Buckley, J., *Fuzzy probability and statistics*, Springer-Verlag, Berlin, 2006.
4. J. Fodor and G. Bardossy, *Uncertainty and risk: Mathematical concepts and some geological applications*, 2nd Serbian-Hungarian Joint Symposium (SISY 2004), 2004.
5. G. Kuznetsov, V., *Litologiya. osadochnye gornye porody i ih izuchenie [lithology. sedimentary rocks and their study]*, Moscow: “Nedra-Biznescentr”, 2007.
6. M. Navara, *Probability theory of fuzzy events*, EUSFLAT-LFA Conf., 2005, pp. 25–329.
7. I. Orlov, A., *Teoriya nechetkih mnozhestv – chast' teorii veroyatnostej [fuzzy set theory is a part of probability theory]*, Nauchnyj zhurnal KubGAU [Scientific journal KubSAU] **92(08)** (2013), 51–62.
8. M. Poleshchuk, O., *Nechetko-veroyatnostnye prostranstva i veroyatnosti nechetkih sobytij [fuzzy probability spaces and probabilities of fuzzy events]*, Lesnoj vestnik [Forest Bulletin] **4** (2014), 162–169.
9. M. Sinitsyna V., *Avtomatizirovannyj analiz struktury mikrotekstury shlifa peschanika [automated analysis of the structure of the microtexture of a thin section of sandstone]*, Aktual'nye problemy nefti i gaza [Actual problems of oil and gas], 2019.
10. S. Torno, J. Torano, M. Menendez, M. Gent, and J. Velasco, *Mathematical and fuzzy logic models in prediction of geological and geomechanical properties of rock mass by excavation data on underground works*, Journal of Civil Engineering and Management **17** (2011), no. 2, 197–206.
11. G. Wang, Y. Xu, and S. Qin, *Basic fuzzy event space and probability distribution of probability fuzzy space*, Mathematics **7,542** (2019).
12. J. Wang, W. Cui, and J. Ye, *Slope stability evaluation using tangent similarity measure of fuzzy cube sets*, Journal of Soft Computing in Civil Engineering **3** (2019), no. 1, 27–35.
13. A. Zadeh, L., *Probability measures of fuzzy sets*, J. Math. Anal. Appl. **23** (1968), 421–427.
14. Y. Zhang, M. Oldenburg, C., and S. Finsterle, *Percolation-theory and fuzzy rule-based probability estimation of fault leakage at geologic carbon sequestration sites*, Environmental Earth Sciences **59 (7)** (2010).

V.Yu. Itkin: Department of Applied Mathematics and Computer Modelling, National University of Oil and Gas “Gubkin University”, 119991, 65 Leninsky Prospekt, Moscow, Russia

E-mail address: `itkin.v@gubkin.ru`

V.M. Sinitsyna: Department of Applied Mathematics and Computer Modelling, National University of Oil and Gas “Gubkin University”, 119991, 65 Leninsky Prospekt, Moscow, Russia

E-mail address: `v-sin@yandex.ru`