


Kryanev A.V., Semenov S.S.

ON THE ISSUE OF QUALITY AND RELIABILITY OF EXPERT JUDGMENTS IN DETERMINING THE ENGINEERING LEVEL OF COMPLEX SYSTEMS

This paper has shown the topicality of using a method of expert judgments for developing complex technical systems (CTS). And in reference with this, the paper has considered the issues of judgments of a required number of experts in conducting peer reviews, their expert knowledge, data on the minimum size of an expert group and methodical approach to determine the level of possible errors of experts to provide judgment of an expert group, including expert assessments on the basis of past knowledge, as well as the possibility of an aggregate consideration of opinions of individual experts involved in the assessment of a case study.

Keywords: *complex technical systems, method of expert judgments, accuracy of an expert judgment, expert knowledge, number of experts as part of a group, weight factor, dispersion of expert judgment errors.*

The topicality of expert judgment method

Expert judgment methods based on the expertise and experiences in carrying out research and development (R&D), study of information on scientific and technical literature, periodicals and patents in a specific area combined with modern mathematical methods of knowledge are an important tool for determining general progress trends, parameters, and technical level of complex technical systems (CTS) [1,2].

Generally, the method of expert judgments is used in the study of entities and problems, the analysis of which is either fully or partially cannot be formalized, i.e. where it is difficult to develop an adequate model. A procedure for conducting expert judgment methods is comprehensively enough described in scientific literature, for example in [1], with the consideration of the issues related to forming expert groups, expert knowledge, forms of work with experts, and the formation of questionnaires and tables of expert examinations, and analysis and expert judgments' processing, also with coordination of opinions of individual experts taken into account. The determinative factors in conducting expert judgments defining the judgment reliability and its accuracy is the selection of experts and their expertise.

Judgment accuracy and the number of experts

It is very important to define the appropriate numerical composition of an expert team. With an insufficient number of experts, the results of their work will not be reliable. A large team of qualified experts is difficult to form and organize its work.

The method of CTS technical level assessment has been developed with participation of experts, the main provisions of which are given in [3-5]. In this case, for the assessment of CTS technical level, experts are usually involved to generate indicators and to define the weight (importance) of estimated figures and ranks of objects under study. An important factor for a researcher is reasonable selection of experts.

To form an expert working team is the important stage in the process of expert judgment, and it is necessary, in case of need to allocate sub-teams of experts in a particular area, and the number of experts in the team.

Selection of expert working teams is carried out in the following succession:

1. A list of questions that need an expert opinion is prepared.
2. A list of experts who together can provide answers for the whole set of questions is formed.
3. A list of questions is sent to each expert, in order to find out whether he can or cannot competently participate in the examination of the question under consideration.
4. Next, it is necessary to determine spending of time and resources that could give an answer for each question, with a minimum of time and resources to carry out the survey.

The expert must meet the following requirements:

- expert judgments should be stable over time during the time of examination;
- the expert should be qualified in a given field of knowledge, i.e. he should be a respected authority in field of the development of new products of this type when evaluating the technical level of CTS.

It is advisable to involve in expert judgments such experts, whose judgments consistently differ from true assessments by downward bias or upward bias, i.e. the error is of systematic nature. In this case, you can identify a bias and make corrections for it.

In the binomial model of choice, the larger is the sample size, the more accurate is the estimation of a characteristic. It is believed that a minimum of sample sizes commonly used in marketing research makes up 100. [6] As the cost of research grows as the linear function of a sample size, and the accuracy increases as the square root of this size, then the upper limit of the sample size is also selected based on economic considerations.

The number of experts should be large enough, so that individual opinions should not be unduly significant. However, the sudden increase in the number in the group reduces their level of competence, which can significantly reduce the accuracy of expert judgment.

For the calculation of the number of experts, a well known ratio can be used, which is applied to calculate the error of observations [7]

$$N = t_p^2 / \varepsilon_1^2,$$

where N is the number of experts in the team,
$\varepsilon_1 = \varepsilon / S$ is the relative error margin of expert judgment,
S is a root-mean-square deviation of the distribution of assessments ratings of any quantity,
t_p is Student's coefficient, which determines the width of the confidence interval and the dependence on the probability estimates P (t_p is a tabulated value).

The minimum acceptable number of experts in the team *N* (Table 1) can be determined depending on the specified accuracy of expert judgment and selected probability estimates.

Table 1. The minimum acceptable number of experts in the team

ε_1	Probability estimates *P*							
	0,99	0,95	0,90	0,85	0,80	0,75	0,70	0,65
0,5	26	15	11	8	7	5	4	4
0,3	74	43	31	23	19	15	12	10

It is empirically established that the number of experts of 13-15 persons can be considered as a sufficiently representative group for examination of CTS.

Proof of this is the dependence of the accuracy and reliability of the assessment of the event occurrence term with a number of experts in the team *N* (Fig. 1).

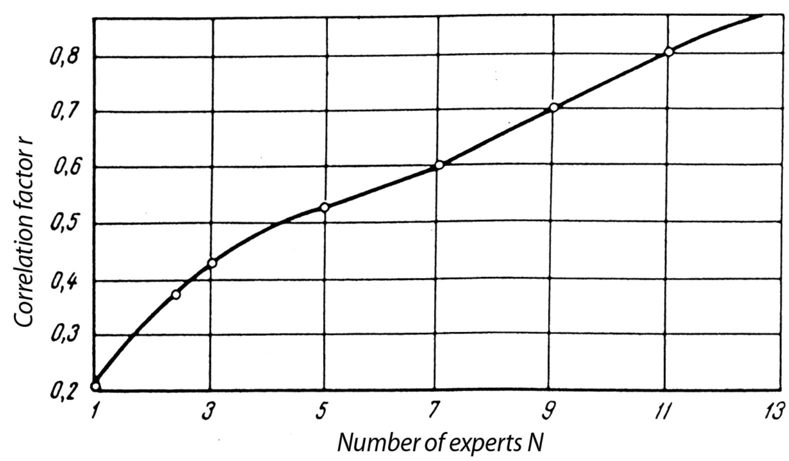

Fig. 1. Relationship of accuracy and reliability of results' assessment of the event occurrence term with a number of experts in the team *N*

Definition of an acceptable quantity of an expert team

In connection with mentioned above, it is of interest to give an example presented in [1]. Suppose that there are 100 professionals on the matter under discussion. Due to the organizational difficulties, we can create a team of experts consisting of 50 persons. For simplicity of calculation, we assume that an expert is such a professional, who has experience on the subject of at least 10 years. Then the original problem is formulated as follows: it is necessary to determine the 50-percent sample of 100 professionals with a number of experts with experience of at least 10 years with the probability of 0.9545.

Suppose that the proportion of professionals with experience of 10 years and more, from some a priori specified hypotheses is equal to 0.6 and professionals with experience less than 10 years respectively is equal to 0.4. Representativeness error can be calculated in accordance with the theorem of Bernoulli by the following formula

$$M_g = t\sqrt{\frac{rg}{n}},$$

where M_g is the error of representativeness;
t is the confidence factor;
r is the portion of sample units with the presence of the given criterion (in this example r = 0,6);
g *is* the portion of sample units with no given criterion (g = 0,4). At the specified probability $P = 0.9545$, coefficient $t = 2$.

Then

$$M_g = 2\sqrt{\frac{0{,}6\cdot 0{,}4}{50}}.$$

Thus, in the original team of professionals the part of experts with experience not less than 10 years is 0,6 ± 0,138, or it will be within the limits from 46.2 to 73.8%.

Practice of prediction tends to the minimally possible number of experts in the team. Reducing the number of experts lower than a certain limit is equivalent to decrease the accuracy of a sample. In such a situation, the problem of establishing the quantity number of an expert team with defined values of decreasing the accuracy of a sample arises. Let a sample reduction of 10% in the previous example be given, which makes up 0.0138 of the calculated error of representativeness. Then the resulting error will make up $\Delta_g = \pm\,(0{,}138 + 0{,}0138) = \pm\,0{,}1518$.

The sample size is calculated by the formula

$$ng = (t^2\, r\, g\,)/\,\Delta^2_{\,g} = (4 \cdot 0{,}6 \cdot 0{,}4)\;/\;0{,}0225 = 36 \text{ experts.}$$

Thus, in a team of 100 professionals there are 36 experts with probability $P = 0.9545$. As a result of studies to determine the quantity of an expert team, given in [8], the dependence of $\beta/\alpha = f(k)$ (β is variation, α is the width of confidence interval, k is the number of experts) for the selected confidence probability p (Fig. 2) has been obtained.

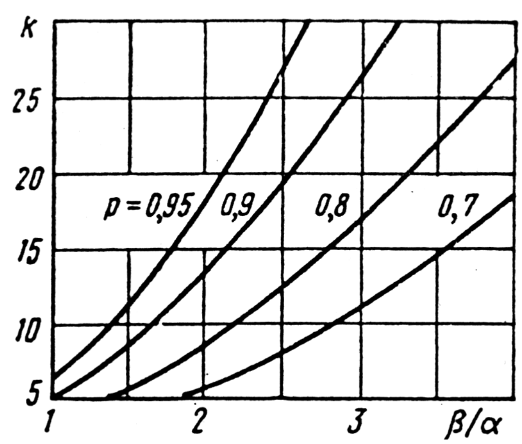

Fig. 2. The diagram for determining the quantity of experts k in terms of the ratio β/α

Suppose that we want to create a team of experts to evaluate the technical level of CTS, and it is desirable that the confidence interval of expert judgment of a non-metric criterion with the probability 0.8 was no more than ± 10% of its average value. Based on previous experience, it is known that those to be included into the team of professionals in problems of this kind gave the value of β, which does not exceed 0.25.

Thus, as input data we have: $\alpha = 0{,}1$; $\beta = 0{,}25$, i.e. $\beta / \alpha = 2{,}5$; $p = 1$. For these initial data, from the graph (Fig. 2) we find $k = 12$.

Thus, we conclude that the best solution will be to organize an expert team consisting of 10-12 persons. In this case, it is assumed that all the experts are familiar with the assigned problem and ready to answer all raised questions. However, sometimes when assessing the technical level of CTS, questions raised for experts have a so varied nature (e.g., technical, operational, economic, etc.), that it is difficult to find a professional, who is able to deal adequately with the total spectrum of issues. In such cases, the expert team should involve along with experts on all issues (limited number) experts in a narrow field who can only answer certain questions. As a result, a full team of experts is formed. But the question arises whether the assessment obtained by the given expert staff is reliable? Indeed, in this case, the expert quality does not apply to the entire set of questions posed to experts.

Competence of experts

Assessment of experts' competence is not less important than the determination of the minimum quantity of an expert team. Obviously, depending on the nature and type of products the coefficient of competence of the same expert may vary from zero (completely incompetent) to one (the highest value of the weighting factor).

It should be noted that at present the objective methods of assessment of experts are difficult to realize practically, and sufficiently subjective methods such as self-appraisal, mutual appraisal and appraisal based on documents of qualification are mainly used.

In the first approximation, we can evaluate the quality of an expert, using a formal approach to the assessment of his professional competence. [7] First, professional competence is determined by the level of academic qualifications of an expert K_K, which can be appraised approximately from Table 2.

Secondly, the competence of an expert is determined by the structure of arguments, which served as the basis for appraisal of his reasoning factor (factor K_a), and the degree of his knowledge of a problem under study (factor K_s).

Factor of argumentation K_a can be obtained, in particular, by adding the appropriate numerical values in the table (Table 3) marked by an expert as self-appraisal procedure (the list of sources of reasoning is given in the questionnaire).

Table 2. Approximate level of scientific qualifications of an expert

Position	Qualification factor values K_K			
	Without a degree	PhD	Doctor of engineering	Corresponding member, Academician
Senior researcher	0,15	0,225	0,30	0,50
Head of Laboratory, team leader, associate professor	0,20	0,30	0,40	0,60
Head of department, Deputy head of department, head of a chair	0,25	0,375	0,50	0,75
Head of a complex, Deputy head of a complex	0,30	0,60	0,60	0,90
Director, Deputy director, Problem research manager	0,40	0,60	0,80	1,00

Table 3. Appraisal of expert opinion reasoning

Sources of reasoning	The degree of source influence		
	high	average	low
Field experience	0,40	0,30	0,20
Conducted theoretical analysis	0,20	0,15	0,10
Accounting trends identified at recent conferences and workshops	0,10	0,10	0,05
Generalization of the works of local authors	0,10	0,05	0,05
Generalization of the works of foreign authors	0,05	0,05	0,05
Personal knowledge of the state of affairs abroad	0,05	0,05	0,05
Intuition	0,05	0,05	0,05
Reasoning factor K_a	1,00	0,80	0,50

It should be noted that the numerical judgments in Table 3 are approximate, but it is important to emphasize that the total reasoning factor K_a should not exceed 1. The degree of influence on an expert opinion of all the listed sources of reasoning ($K_a = 1$ – high, $K_a = 0.8$ – average, $K_a = 0.5$ – low) is determined, as already noted, by an expert himself.

The degree of expert knowledge of a problem is appraised similarly. In the questionnaire, an expert fills in score of self-appraisal (e.g., from 1 to 10), and the maximum score (10) corresponds to the knowledge at the level of authorship (co-author) in development of specific approaches to the problem solution, and the minimum score (0) is a complete lack of problem knowledge. Next, the normalization of the scoring is carried out, i.e. by multiplication by 0.1 the factor $K_з$, is brought within the range from 0 to 1.

Versatility indicator of competency of the i-th expert can be calculated under the following formula

$$K_i = (K_к + K_a + K_з) / 3.$$

The value K_i is in the range from 0.05 to 1.0.

Typically, the factor of professional competence of an expert is taken into account as a multiplier in appropriate assessment, i.e. as it serves as a factor of importance (weight) of a given expert opinion (for example, in assessing values of the weight function in the matrices of levels of consumer properties and total costs).

In [8] the degree of competence defines the credibility factor, which means the number that shows what weight should be included into the statistical analysis of the quantitative evaluation of a given expert. The credibility factor is expressed in normalized form, and factor coefficient $\mu = 1$ is attributed to maximum credibility, and less competent experts have $\mu \angle 1$, incompetent expert should get $\mu = 0$.

In order to select a reasonable method for determining the credibility factors of experts according to the results of the expertise itself, the study [8] offers the approach to the formation of this factor.

1. The credibility factor should be in the range 0-1.
2. The credibility factor should be a monotonic function of the difference between the individual and the average assessment of the system criteria under study.
3. For the same difference between the individual and the average assessment the credibility factor shall be the smaller, the smaller is the variance of the averaged value of criterion.

4. At the concurrency of the individual and the averaged assessment the credibility factor should turn into 1, that is, possess its maximum value.

5. The credibility factor should be determined using expert judgments of criterion under study for all the competing systems in a given problem of choice.

Peculiarities of propositions 3 and 5 are presented in detail in study [8].

We have developed an approach that offers to use the information about the levels of potential errors of experts and takes into account this information in obtaining an aggregate assessment of an expert team [9]. Refinement of the average value with additional information embedded in the analyzed sample of data can be done based on the Bayesian scheme [10]. The Bayesian approach is widely used as a tool of the theory of probability in various fields. To get a consolidated judgment of an expert team, the method is offered that takes into account the "distances" of experts' judgments to the aggregate one based on which the experts are assigned different weights.

Thus, for n criteria and m experts we have $m \times n$ judgments, which can be represented as a matrix:

$$X = \begin{bmatrix} x_1^1 & \dots & x_n^1 \\ \dots & \dots & \dots \\ x_1^m & \dots & x_n^m \end{bmatrix}, \quad (1)$$

where $x_j^{(i)}$ is the weight of the criterion j according to the opinion of the expert i.

Then we calculate the arithmetical mean of judgment values for each criterion

$$\overline{\overrightarrow{x^{(A)}}} = \frac{1}{m}\sum_{i=1}^{m}\overrightarrow{x^{(i)}} = \frac{1}{m}\left(\sum_{i=1}^{m}x_1^{(i)},\dots,\frac{1}{m}\sum_{i=1}^{m}x_n^{(i)}\right) = (\overline{x_1^{(A)}},\dots,\overline{x_n^{(A)}}), \quad (2)$$

where $\overline{x_j^{(A)}}$ is the average value by criterion j.

Further refinement of the obtained values of weight factors of criteria significance are based on consideration of different level of errors in judgments of different experts, which will be considered in the calculation of the consolidated judgments.

We shall introduce the random variable whose realizations are equal to deviations of the significance of experts' criteria from the true value. It is important to note that in this case, arithmetical mean value of experts' judgments is taken as true value.

Then, for every expert we have the estimate of variance of the introduced random variable:

$$\sigma^{(i)2} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j^{(i)} - x_j^{(A)})^2. \quad (3)$$

Then we shall calculate the weight factor for each expert:

$$w^{(i)} = \frac{1}{\sigma^{(i)2}} / \sum_{i=1}^{m}\frac{1}{\sigma^{(i)2}}. \quad (4)$$

It is important to note that in this case we take the assumption of stochastic independence of experts' judgments. This assumption is based on the fact that experts do not report scores given to their colleagues, which provides independent experts' judgments as random variables.

After obtaining weight factors (4) it is necessary to calculate the refined judgments of criteria significance as weighted averages taking into account different levels of expert judgments' errors:

$$\overline{x_j^{(B)}} = \sum_{i=1}^{m} w_i \, x_j^{(i)} . \tag{5}$$

As a result, for each criterion j the value $\overline{x_j^{(B)}}$ will be found, which corresponds to the consolidated judgments of all experts in view of individual errors of experts. This value is the weighted average value of experts' judgments based on their expert competence.

Another reliable scheme for determining the competence of each expert expressed in the definition of weights of their judgments is the estimation of errors variance of experts' judgments in the past. This scheme is applicable if the matrix of experts' judgments is known, and implemented values of the predicted characteristic are also known and they can be represented as a matrix

$$\begin{pmatrix} X_{11}, & X_{12}, ..., & X_{1m}, & Y_1, \\ X_{21}, & X_{22}, ..., & X_{2m}, & Y_2, \\ ..., & ..., \; ..., & ..., & ..., \\ X_{N1}, & X_{N2}, ..., & X_{Nm}, & Y_N, \end{pmatrix}, \tag{6}$$

where X_{ij} is the numerical value of the judgment of the j-th expert in the i-th judgment carried out in the past, Y_i is the implemented value of the characteristic under judgment.

Based on the known values of the matrix elements (6), it is possible to estimate a covariance matrix of relative errors of experts' judgments, using the following formula

$$\mathrm{cov}(X_i, X_i') = \frac{1}{N} \Sigma_{k=1}^{N} \left(\frac{X_{ki}}{Y_k} - 1 \right) \cdot \left(\frac{X_{ki}'}{Y_k} - 1 \right), i, i' = 1, ..., m . \tag{7}$$

In particular, for $i = i'$ the formula (7) gives variances of relative errors of experts' judgments.

Then the optimal final assessment of the characteristic under experts' judgments with the aggregate of m experts is given by the following equality

$$X_{opt} = \frac{\left(K_{\overline{X}}^{-1} \overline{1} . \overline{X} \right)}{\left(K_{\overline{X}}^{-1} \overline{1} . \overline{1} \right)}, \tag{8}$$

where $K_{\overline{X}}^{-1}$ is the inverse of the covariance matrix $K_{\overline{X}}$ of the relative errors of experts' judgments ((i, i ') is an element of this matrix is equal to the covariance (7)), $\overline{1} = (1,1,\ldots,1)^T$ – m is the N-dimensional vector, $\overline{X} = (X_1, ..., X_m)^T$ is the vector of consolidated experts' judgments.

The variance of the relative error of the optimal estimate (8) is given by the following equality

$$\sigma^2(X_{opt}) = \frac{1}{(K_{\overline{X}}^{-1}\overline{1}.\overline{1})}. \qquad (9)$$

Optimality of estimate (8) indicates that the variance of relative error of judgment of any of *m* experts is more than the variance (9).

Thus, the proposed scheme of consolidating judgments of several experts, having optimal properties, figuratively speaking, "squeeze" all the information from judgments of individual experts and provide a summary unifying judgment with the least error[1].

In assessing the technical level of CTS, diverse questions are raised for experts (for example, to assess the significance of estimated figures such as "technical", "technological", "operating", "economic", etc.), and therefore it is difficult to find an expert who is able to understand competently all issues. In such cases, the expert team should include along with a broad specialists, highly qualified professionals in narrow fields. The number of such professionals is determined by their level of professionalism and solved with help of equality (9).

The geometric mean value is taken for aggregation of expert opinions calculated by the following equation [11]

$$a_{ij}^A = \sqrt[n]{a_{ij}^1 a_{ij}^2 ... a_{ij}^n}, \qquad (10)$$

where a_{ij}^A is the aggregate judgment of the elements belonging to the *i*-th row and *j*-th column of the matrix of pairwise comparisons; *n* is the number of matrix of pairwise comparisons, each of which is drawn up by one expert.

The consistency of criterion (10) becomes apparent when two equivalent experts indicate *a* and *1/a* respectively when comparing objects, and in calculating the aggregate judgment it gives 1 and testifies to the equivalence of compared objects.

Conclusions

Based on the carried out analytical studies on the problem of assessing the accuracy and reliability of experts' judgments in the evaluation of technical level of CTS:

1. The procedure for the formation of an expert team and the requirements to experts has been offered.
2. As sufficient for practical calculations, composition of an expert team with a number of expert equal to not less than 10-12 persons at the confidence level of 0.8 has been defined.
3. The approaches and alternatives to determine the competence factor of an expert, which takes into account factors of scientific skills, reasoning of expert opinion and knowledge of the estimated problem have been considered.
4. The methodical approach to the evaluation of possible errors of experts in obtaining consolidated judgment of experts has been offered.

[1] When combining assessments of several experts it is often orientated by the statement that the number of experts should be large enough. The above presented schemes show that it is often sufficient to have the judgment of only one highly qualified expert, in order that the estimation error should be below the acceptable level.

References

1. **Yampolsky S.M., Lisichkin V.A.** Prediction of scientific-technical progress – Moscow: Economics, 1974. – 207 p.

2. **Scheverov D.N.** Designing of unmanned aerial vehicles. – Mechanical engineering, 1978. – 264 p.

3. **Semenov S.S.** Assessment of the technical level of guidance systems for controlled aircraft weapon of the class “air-to-surface” // The bulletin of computer and information technology. – 2006. – # 8 – p. 7-11; # 9 – p. 13-23; # 10 – p. 12-18.

4. **Semenov S.S., Shcherbinin V.V.** Method of assessing engineering level of guidance systems of robot bombs / Proceedings of the Fourth theoretical and practical conference “Advanced Systems and Control Problems” and of the first youth school workshop “Managing and processing information in technical systems”. Taganrog. TFYU, 2009 – 291 p. – P. 160-167.

5. **Semenov S.S., Shcherbinin V.V.** Method of assessing engineering level of guidance systems of robot bombs // Issues of enginery. Line. 9. Special control systems, follower drives and components .. – Moscow: FSUE “STC”Informtekhnika”, 2010. – Vol. 1 (242) – 2 (243). – 108. – S. 29-32.

6. **Orlov A.I.** Econometrics. Rostov-on-Don: Phoenix, 2009. – 572 p.

7. **Bobrovnikov G.n., Klebanov A.I.** Integrated prediction of a new technology development. – Moscow: 1989. – 205 p.

8. **Brahman T.P.** Multicriteria choice and selection of an alternative in engineering. – Moscow, Radio and Communication, 1984. – 288 p.

9. **Kryanev A.V., Tikhomirov A.N., Sidorenko E.V.** Group expertize of innovation projects using Bayesian methods / / Economics and Mathematical Methods. – 2013. Tom 49. – #2. – P. 134-139

10. **Kryanev A.V., Lukin G.V., Udumyan D.K.** Metric analysis and data processing. – M. Physmathlit, 2012. – 308 p.

11. **Andreychikov A.V., Andreichikov O.N.** System analysis and synthesis of management decisions in innovation: Math, heuristics and intelligent methods of system analysis and synthesis in innovation. – Moscow: Publishing house “LIBROKOM”, 2013. – 304 p.