APPLICATION OF FUZZY MODELING TO PREDICT THE DISEASE OF STAFF FROM EXPOSURE TO WORKING CONDITIONS

Klimova I. V., Smirnov Yu. G.
Department of Informatics, Computer Technologies and Engineering Graphics - Ukhta State Technical University, the Russian Federation

bgd4@mail.ru, ysmirnov@ugtu.net

Abstract: *A fuzzy model for determining the morbidity rate of employees of a refinery with diseases of the respiratory organs is analyzed on the basis of an analysis of the concentrations of pollutants in all occupational environments using a mathematical apparatus of fuzzy sets. The results of visualization of the developed fuzzy model in the MATLAB Fuzzy Logic Toolbox medium are presented.*

Keywords: FUZZY MODELING, MODEL, WORKING CONDITIONS, HARMFUL PRODUCTION FACTORS, MORBIDITY, STAFF

1. Introduction

Fuzzy modeling is the most promising direction for scientific research in the field of analysis, forecasting and modeling of various processes. This is especially important for assessing occupational risks for the health of personnel, where there is insufficient data on the connection of certain diseases with working conditions. The existing assessment of working conditions makes it possible to determine the "verbal" level of risk on the basis of the established class of working conditions. At the same time, the range of production factors under consideration is constantly narrowing, excluding even the receipt of additional payments for harmful working conditions. Thus, with the invariance of working conditions, the class of working conditions is reduced only by changing the methodology by which labor conditions are evaluated.

The purpose of this work is to assess the applicability of existing models and methods of fuzzy logic to modeling the impact of harmful substances on the health of personnel of "RN - Komsomolsk Oil Refinery".

The object of the study is the personnel of the plant exposed to harmful substances in three environments for the period from 2004 to 2010. The subject of the study is inhalation non-carcinogenic risks for the health of the personnel of the oil refinery.

2. Problem discussion

Consider the problem of constructing the dependence of the morbidity of diseases of the respiratory organs of the personnel of the oil refinery on the indices of the non-carcinogenic hazard of chemicals and suspended substances. The nosological group of "respiratory diseases" was chosen on the basis of an earlier analysis of the impact substances and the peculiarities of their effect in three environments: industrial-technological, industrial and urban [1].

The production technological environment is the most polluted - it is the territory of technological workshops and installations. The production environment is a less polluted environment within administrative buildings, as well as the rest of the plant's territory. The urban environment in turn is a combination of household and environmental environments in the city of Komsomolsk-on-Amur. Such a division into environments is made with the aim of clarifying the concentrations of harmful substances, and, accordingly, a more detailed assessment of the health risks of workers [1].

A total of 66 substances participated in the analysis. Figure 1 shows the quantitative ratio of substances affecting personnel. Of them, substances affecting the respiratory system - 30 items, 7 has a primary effect on the respiratory system (Figure 2) [2].

All harmful substances that affect the respiratory system can be divided into two groups: due to their specific effects: chemical substances and suspended substances. Therefore, the model of the dependence of the morbidity of personnel with respiratory diseases on two parameters will be constructed: the index of non-carcinogenic hazard for chemical substances and the index of non-carcinogenic hazard for suspended substances.

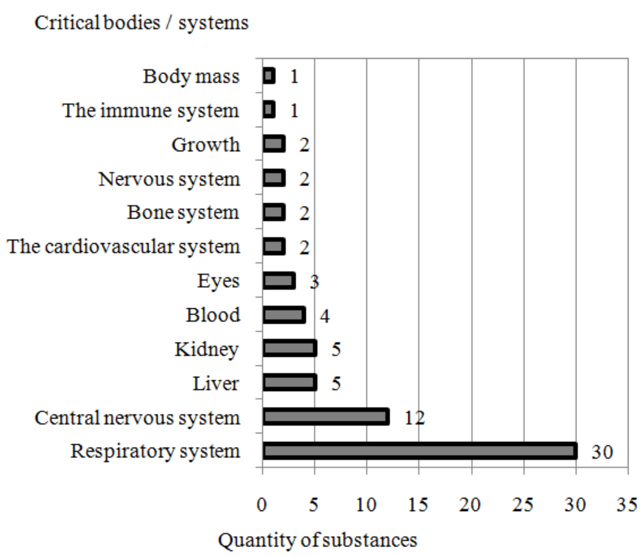

Fig. 1. *Ratio of substances affecting critical bodies / systems*

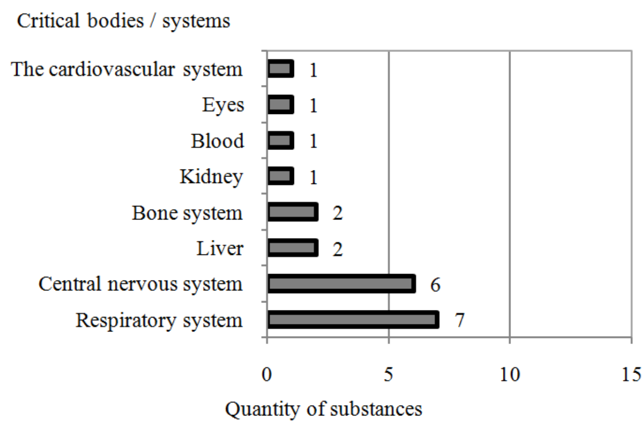

Fig. 2. *Ratio of substances according to their primary effect on critical organs / systems*

Indices of non-carcinogenic hazards are quantitative estimates of the amount of harmful substances affecting the worker's body throughout the day in three environments and are calculated first for each substance (1) and then for the group of effects on the particular organ or system as a whole (2):

$$(1)\quad HQi = AC / RfC,$$

$$(2)\quad HI = \Sigma\, HQi,$$

where *HQi* are the hazard ratios for the individual components of the mixture of agents; *HI* - index of non-carcinogenic hazard for a critical organ or system; *AC* - average concentration of substance, mg/m^3; *RfC* - reference (safe) concentration, mg/m^3 [1].

It is worth noting that the average concentration of each substance involved in the calculation was averaged over 24 hours, depending on each type of medium, the concentration of the substance in it, and the time of exposure to the worker.

After the *HI* indices for respiratory organs were calculated at each workplace, the data were averaged over 29 occupational groups. The main criteria for the formation of groups of personnel: belonging to the facility or shop and the value of the received non-carcinogenic risk *HI*.

The minimum values of the indices were $HI_{min\ chem}$ = 23.87 and $HI_{min\ dust}$ = 2.56, which exceeds the permissible value HI_{norm} = 1. The maximum values of the indices were $HI_{max\ chem}$ = 1621.58 and $HI_{max\ dust}$ = 5.01.

Next, consider such an output parameter as the state of health of personnel, which is a reflection of a complex set of phenomena in the environment. The process of its formation is influenced by a number of industrial, socio-economic, as well as biological, anthropogenic, natural climatic and other factors that together determine the ecological environment in which the person is during the day, food and water. Most of the harmful substances a person receives with inhaled air, this is about 80% of all intake doses.

Anticipate the corresponding incidence rate of personnel with respiratory diseases is quite difficult, especially with the seasonality of this phenomenon. In this connection, it is expedient to use linguistic variables, that is, variables whose values are not numbers, but words in natural or formal language.

The data on the morbidity of the personnel of the plant on the nosological form of "respiratory disease" from 2004 to 2010 were broken down into the allocated 29 occupational groups, and then averaged over 7 years and included the number of cases per 1000 workers and the number of days of incapacity for work 1000 working. Further, data are used only for the number of cases of diseases: the minimum value is 125 diseases, the maximum is 495.

Comparison of calculated *HI* indices and morbidity leads to the receipt of 29 points, which are difficult to describe with the help of the equations of dependence, but nevertheless, a directly proportional relationship is clearly visible: the greater the hazard index, the greater the response.

For the transition to fuzzy logic, an additional conversion of the input parameters (*HI* indices) was performed. Since all indices exceeded the allowable value, an attempt was made to select a so-called "acceptable" value that would guarantee minimal deviations in the state of health.

For this purpose, an additional criterion "severity of disease progression" was introduced and a corresponding group was chosen whose indices are accepted for an "acceptable" new standard. This made it possible to obtain new ranges of values of $HI_{chem} \in [0.12, 8.05]$ and $HI_{dust} \in [1{,}0, 1{,}96]$.

3. Methodology

The modeling process can be presented in large form in the following sequence of actions:

1) awareness of the problem;
2) highlight the main factors that determine the problem, which should serve as output parameters of the model;
3) highlighting the defining input variables of the model;
4) the actual development of a mathematical model;
5) identification of the model (parametric or structural-parametric);
6) conducting numerical experiments with the model and, if necessary, statistical processing of the obtained data;
7) determining the composition of the quality parameters that characterize the problem being solved (using simulation data and a priori information);
8) formalization of particular quality criteria based on quality parameters;
9) determination of parameters characterizing the relative importance of particular criteria for solving a common problem;
10) formalization of the generalized quality criterion for solving a problem on the basis of aggregation of particular criteria, taking into account their relative importance;
11) solving the problem of choosing the best alternative or multi-criteria optimization, depending on the type of problem being solved.

In essence, it is some detail of the generally accepted scheme: the formulation of the problem → model building and identification → optimization.

The development of fuzzy models of sanitary and toxicological safety of the personnel of industrial enterprises makes it possible to obtain a numerical estimate of occupational risk. The mathematical apparatus of fuzzy logic is usually used in those cases when the available quantitative information is insufficient, or it is not complete enough to obtain reliable statistically significant conclusions [3-7].

Along with classical analytical methods, it is advisable to use the fuzzy set device implemented in particular in the MATLAB computer simulation system [8], which allows developing a fuzzy-multiple model for the estimation, analysis and visualization of professional risk indicators.

Since there is a need to take into account the multitude of indicators that are dissimilar in physical nature and dimensions, it is advisable to bring them to a dimensionless form by rationing, for example, as follows:

$$(3)\quad s = \frac{s_i - s_{min}}{s_{max} - s_{min}},$$

where s_i is a normed index; s_{max}, s_{min} - the maximum and minimum value of the criterion in the sample according to the normed indicator.

The system of fuzzy inference in the general case includes the following stages:

1. Phasing (reduction to fuzziness). At this stage, the exact set of input data is converted into a fuzzy set, which is determined using membership functions.
2. Construction of the base of rules for fuzzy products.
3. Composition using aggregation methods.
4. Dephasing (reduction to clarity). At the stage of dephasing, the fuzzy system's executive module, on the basis of many fuzzy conclusions, forms an unambiguous decision with respect to the input variables.

Let us consider in more detail the initial stage on which phasing is carried out. Denote by d the input variable "suspended matter", which reflects the dustiness of the company's air environment. The corresponding term-set will be denoted by:

T1 = {low, medium, high} = {*D1*, *D2*, *D3*}.

The second input variable x - "chemical substances" - reflects the chemical contamination of the air environment. It corresponds to the analogous term set:

T2 = {low, medium, high} = {*X1*, *X2*, *X3*}.

The output variable y (the level of morbidity of workers) is also comparable to the analogous term set:

T3 = {low, medium, high} = {*Y1*, *Y2*, *Y3*}.

The next stage is the construction of a database of rules for fuzzy products. Most often, the Mamdani model is used as a model of fuzzy inference, the feature of which is the fact that its rules of inference contain fuzzy meanings in its consecutive clauses. In our case this is the membership function of the term-set *T3*.

In the chosen notation, we give for example some of these rules:

IF *d* IS *X1* AND *x* IS *D1* THEN *y* IS *Y1*
IF *d* IS *X2* AND *x* IS *D1* THEN *y* IS *Y2*
IF *d* IS *X3* AND *x* IS *D1* THEN *y* IS *Y3*

As a tool for implementing this approach, it is convenient to use the Fuzzy Logic Toolbox extension of the MATLAB computer mathematics environment, which allows creating fuzzy inference and fuzzy classification systems. The main interactive tool of the Fuzzy Logic Toolbox is the FIS inference editor, which contains tools for the functional mapping of input and output variables [8].

4. The discussion of the results

The stages of the simulation are illustrated in Figures 3-7 below.

Figure 3 shows a view of a software window with a schematic model.

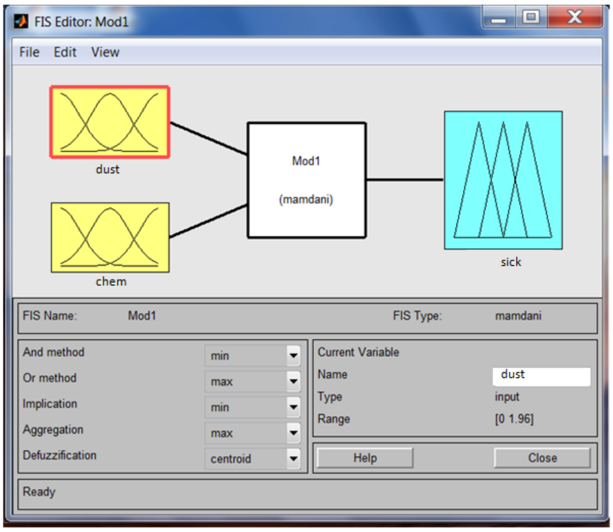

Fig. 3. *Software window view*

Figure 4 shows the graphs of the membership functions for the input terms "dust" (*d*), "chemical" (*x*) and output "sick" (*y*) of linguistic variables.

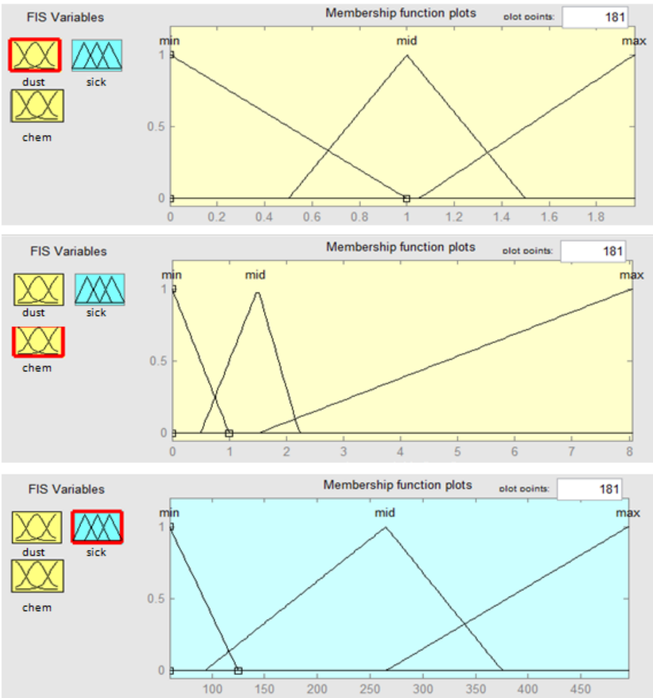

Fig. 4. *Type of membership functions in the FIS editor of the Fuzzy Logic Toolbox extension package: "dust" and "chemical" - for indicators of non-carcinogenic risk for suspended and chemical substances, respectively; "sick" - for the number of cases of respiratory diseases per 1000 workers*

Figure 5 shows a fragment of the window for determining the base of production rules for fuzzy inference.

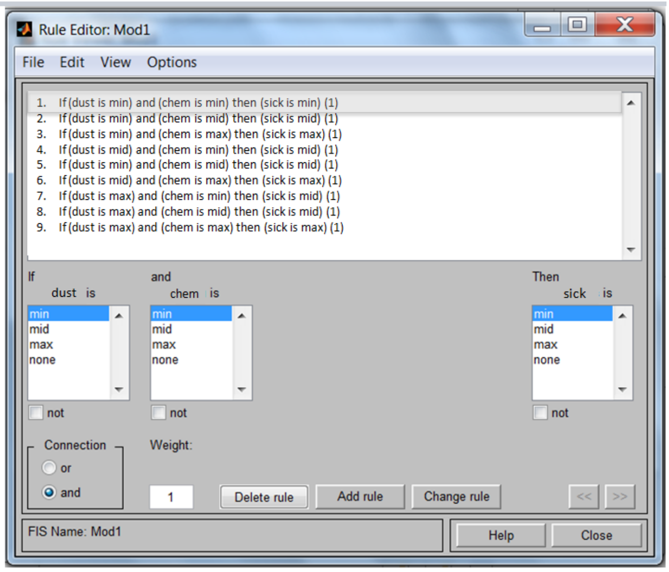

Fig. 5. *The form of the product rules definition window in the FIS editor*

Figure 6 shows the view of the fuzzy output viewer in the Mamdani model for the problem under consideration. Here, the aggregation of fuzzy rules is shown for two input variables, "sick" and "chemical". This uses a logical product, which corresponds to the operation min. Aggregation of implication concerning rules is carried out by logical summation, which corresponds to the operation max.

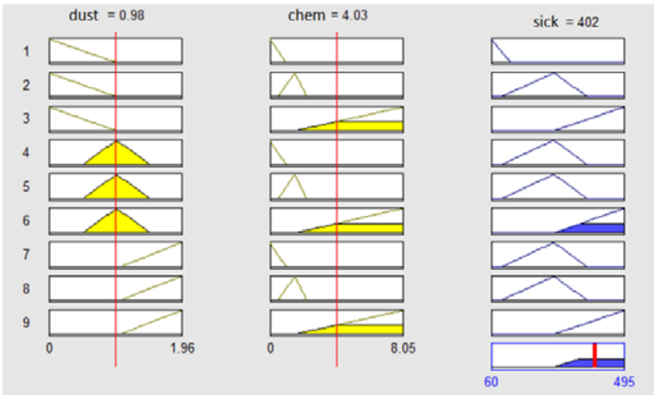

Fig.6. *View of Mamdani's fuzzy output viewer*

Since the max operator is used as an aggregation operator, and the min operator is used as an implication operator, the procedure for obtaining a fuzzy output value is a composition of max-min.

After receiving a fuzzy output (y), it is necessary to go to the phase of dephasing, which has the corresponding clear value y_{out} (4). As a method of dephasing, we used the method of the center of gravity:

$$(4) \quad y_{out} = \sum_{i=1}^{n} y_i\,\mu(y_i) / \sum_{i=1}^{n} \mu(y_i),$$

where $\mu\ (y_i)$ is the membership function of the i-th rule, and n is the number of fuzzy products rules.

Finally, Figure 7 shows the surface of the fuzzy output for the developed fuzzy model. This type serves for a general assessment of the adequacy of the constructed fuzzy model, and also allows analyzing the influence of the values of input variables on the value of the output variable.

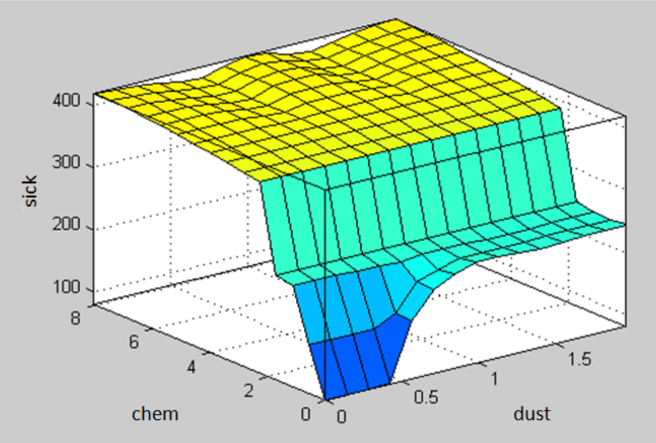

Fig. 7. *The surface of fuzzy output for the developed fuzzy model*

5. Conclusion

1. There are 29 professional groups of plant personnel, the data on which are converted into 2 input parameters *d* (suspended substances), *x* (chemical substances) and 1 output parameter *y* (the incidence of personnel with respiratory diseases).

2. A fuzzy model of the influence of pollutants of two groups on the health of personnel based on the Mamdani model, namely, on the respiratory system, has been developed and mathematically described.

3. A variant of phasing the three identified parameters based on an acceptable risk to staff health is suggested as "low", "medium", "high".

4. In the interactive mode, the fuzzy output system of the solved task is developed and visualized using the graphical tools of the Fuzzy Logic Toolbox extension package of the computer mathematics environment MATLAB.

5. It has been revealed that all personnel of the plant are exposed to a sufficiently large exposure to non-carcinogenic risk, the likelihood of harmful effects on the worker (respiratory diseases) increases in proportion to the increase in the coefficient of non-carcinogenic *HI* hazard.

6. The developed model can be easily supplemented by new indicators of air pollution or other production factors (linguistic variables) and new output parameters (fuzzy inference rules).

6. Literature

1. Assessment of non-carcinogenic risk for professional groups of "RN-Komsomolsk Oil Refinery" / I.V. Afanasyeva, V.V. Anisimov // Scientists notes Komsomolsk-on-Amur State Technical University. Sciences about nature and engineering, 2012. № II-1 (10). P. 89-96.

2. Identification of hazards in the assessment of health risks of personnel of RN-Komsomolsk Oil Refinery / I.V. Afanasyeva, V.V. Anisimov // Materials of the 11th International Scientific and Practical Conference in the Field of Ecology and Safety life activities "Far Eastern Spring - 2011". Komsomolsk-on-Amur : GOUVPO "KnAGTU", 2011. P. 141-149.

3. Diligenskiy N.V., Dymova L.G., Sevastyanov P.V. Fuzzy modeling and multicriteria optimization of production systems in conditions of uncertainty: technology, economics, ecology. M. : "Publishing Machine-Building-1", 2004. 398 p.

4. Mazorchuk M.S., Simonova K.A., Grekov L.D. Application of methods and models of fuzzy logic for modeling of economic processes // Systems of information retrieval, 2007. № 9 (67). P. 159-162.

5. Rogachev A.F., Melikhova E.V. Fuzzy-plural modeling and assessment of ecological safety of agricultural lands with radiation contamination // Global Nuclear Security, 2016. № 1 (18). P. 7-18.

6. Sereda S.N. Assessment of environmental risk through fuzzy models // Mechanical engineering and life safety, 2013. № 3. P. 15-20.

7. Tindova M.G. Fuzzy model of economic assessment of environmental damage // Economics: yesterday, today, tomorrow, 2012. № 3-4. P. 129-139.

8. Leonenkov A.V. Fuzzy modeling in MATLAB and fuzzyTECH. S-Pb. : BHV-Petersburg, 2005. 736 p.