RISK ASSESSMENT OF CROSS COUNTRY PIPELINES USING FUZZY CLUSTERING

E. M. Al-ABD [A], t. EL-SAYED [b], AND M. SHEHADEH [B]

A Department of Mechanical Engineering, Arab Academy for Science & Technology & Maritime Transport
B Department of Marine Engineering, Arab Academy for Science & Technology & Maritime Transport, Alexandria, Egypt.

ABSTRACT: Development of computer-based fuzzy inference system, for risk assessment of pipelines, is always confronted with some difficulties. In conventional Takagi-Sugeno (TS) fuzzy inference systems, the computational complexity increases with the dimensions of the system variables. The number of rules increases exponentially as the number of system variables increases. In such situations, it is not feasible to determine all the rules, in this paper we propose the use of fuzzy clustering methods, in which the fuzzy rules are resulted from measured data. In this paper the indexing pipeline risk assessment methodology is integrated with subtractive clustering fuzzy logic to deal with the uncertainty of the real world conditions. From the pipeline risk assessment indexing method basic rules and scores, a hypothetical data is created to construct the subtractive clustering fuzzy model. A case study for a cross country oil pipeline in Egypt is used to demonstrate the proposed methodology.

Keywords: Pipeline risk assessment, TS fuzzy model, Fuzzy logic, Subtractive clustering.

INTRODUCTION

Pipelines are considered to be the safest, cheapest, most efficient and reliable mean of flammable substances transportation. Due to the huge volume of substances needed to be transported, pipelines would be the only possible mean for transporting the massive quantities of petroleum. According to The Association of Pipelines report, pipelines accounted for 70 percent of all petroleum transportation covering the period 1990 through 2009 (AOPL, 2012) [1].

Although most of pipelines are located underground and partially isolated from the human interference but they are still subjected to many threats and the product leakage could lead to a harmful environmental impact or human loss. The oil and gas companies are putting the safety and integrity of pipelines into account as a primary goal to avoid any leakage or system failure that may lead to disastrous or heavy financial consequences. The pipe failure can never be fully avoided; however, the overall risk of failure can be reduced to an acceptable level by opting efficient risk management strategies (Anjuman Shahriar, Rehan Sadiq, Solomon Tesfamariam, 2012) [2].

Different risk assessment techniques are used by O&G companies, including hazard and operability (HAZOP) analysis, fault tree analysis, scenario based analysis, and indexing methods (Transmission Pipelines & Land Use, 2009) [3].

Pipeline risk assessment is complex and imprecise due to the lack of information or incomplete data. To deal with this uncertainty in such situations, Fuzzy logic system developed by Zadeh, (1965) [4] can be used as a decision making tool by processing linguistic information of such complex structures where this information is represented as fuzzy sets inputs and the output risk values can be represented as a crisp value or fuzzy sets with associated degree of membership (El Sayed, 2009) [5]. For this reason many researchers have adopted fuzzy logic in risk assessment or in many other applications which have imprecise data. Many approaches were presented by using fuzzy reasoning, such as conventional fuzzy inference system (FIS).

For a conventional (FIS), it is a way of mapping an input space to an output space by using fuzzy logic, for the fuzzy reasoning of data the inference system uses a set of membership functions and rules; the fuzzy IF-THEN rules are implemented by experts so they sometimes may be called as fuzzy expert system. In the design stage of fuzzy inference system, one of the important issues is how to reduce the total number of involved rules and their corresponding computation requirements. In a standard fuzzy systems, the number of rules increases exponentially with the number of input variable increases. Suppose there are n input variables and m membership functions for each variable, then it needs m^n rules to construct a complete fuzzy inference system. As n increases, the rule base will be more difficult to implement. This dimensional problem is called "curse of dimensionality" (Ming-Ling Lee, Hung-Yuan Chung, Faung-Ming Yu, 2002) [6].

Fuzzy inference system based on subtractive clustering technique supposed to be one of the suitable solutions for this dimensional problem, where the fuzzy IF-THEN rules are created from input-output data.

Some researches that are concerned in using conventional fuzzy inference system in its applications can be listed as follows: Markowski & Mannan (2008) [7] developed a fuzzy risk matrix that may be used for emerging fuzzy logic applications in different safety analyses (e.g., LOPA). Markowski & Mannan (2009) [8] integrated fuzzy logic with the classical layer of protection analysis, and applying it in pipeline risk assessment to deal with the fuzziness and imprecision of information. Jamshidi, Yazdani-Chamzini, Yakhchali & Khaleghi (2012) [9] proposed an integrated model of fuzzy logic with relative risk score methodology for pipeline risk assessment, using Mamdani algorithm based on experts' knowledge for modeling the uncertainty involved in the problem. Ratnayake (2014) [10] suggested a fuzzy inference system to minimize the suboptimal prioritizations of functions in the functional failure risk (FFR) analysis using an illustrative tailor-made risk matrix, and calculating risk ranks by the suggested FIS. Sa'idi, Anvaripour, Jaderi & Nabhani (2014) [11] proposed a model for the risk of the process operations in the oil and gas refineries. The fuzzy logic system (FLS) was proposed for risk modeling to overcome the uncertainty of the traditional risk based maintenance (RBM) components. Wang, Zhang & Chen (2013) [12] proposed a hybrid approach of fuzzy set theory and conventional fault tree quantitative analysis to quantify the crude oil tank fire and explosion (COTFE) fault tree in fuzzy environment and evaluate the COTFE occurrence probability. Khalil, Abdou, Mansour, Farag & Ossman (2012) [13] integrated the classical layer of protection analysis (LOPA) risk management with fuzzy logic methodology, creating a cascaded fuzzy LOPA to prevent or limit industrial accidents in natural gas plants. Markowski, Mannan & Bigoszewska (2009) [14] applied the fuzzy sets theory on fault and event tree methods by replacing all their variables with fuzzy numbers and obtaining the outcome of each by using one of the defuzzification methods, this application can be further used in the ‘‘bow-tie’’ approach for accident scenario risk assessment. Yuhua & Datao (2005) [15] proposed a method which combining expert elicitation in the fault tree analysis with fuzzy set theories to evaluate probability of failure events concerning oil & gas transmission pipelines and overcome ambiguity and imprecision of some basic events. Aqlan & Ali (2014) [16] proposed lean manufacturing principles combined with fuzzy bow-tie analysis for effective risk assessment process in the chemical industries & to overcome the uncertainty inherent with the risks from traditional bow-tie analysis. Amindoust, Ahmed, Saghafinia & Bahreininejad (2012) [17] proposed a new ranking method on the basis of fuzzy inference system (FIS) for supplier selection problem to handle the subjectivity of decision makers’ assessments in the management of a sustainable supply chain.

In this study we employed the concept of fuzzy logic in order to assess the risks of a pipeline. A number of models are established for the Index Sum and the Leak Impact Factor of a pipeline section. The performance of the constructed models is evaluated in comparison with the hypothetical calculated data and the best fit model is identified based on the performance evaluation indices, including training root mean square error (Training RMSE), check root mean square error (Check RMSE), and correlation coefficient (R^2).

TRADITIONAL INDEXING METHOD

A subjective scoring tool for assessing pipeline risks based on a combination of statistical failure data and operators experience where the pipeline is sectioned to segments according to factors of population, land type, soil condition, coating condition, age of pipeline or any other factors decided by the evaluator.

This method has number of assumptions; where all hazards are independent and additive, the worst case condition is assigned for the pipeline section, all point values are relative not absolute, the relative importance of each item is based on expert judgments, only risks to the public is considered and no consideration for the pipeline operators or contractors.

Data is gathered from records and operators interviews to establish an index for each category of pipeline failure initiator, (a) third party damage, (b) corrosion, (c) design, (d) incorrect operations. These four indices score the probability and importance of all factors that increase or decrease the risk of a pipeline failure. The indices are then summed to be called Index Sum as shown in Eq. 1, as the index sum score increases the probability of risk decreases and vice versa.

The last portion of the assessment addresses the consequences of a pipeline system failure. This consequence factor is called the leak impact factor which is used to adjust the index sum scores to reflect the consequences of failure where a higher point represents a higher risk. The leak impact factor is the product of the product hazards (acute + chronic), leak volume, receptors, and dispersion factor, as shown in Eq. 2, where the dispersion factor is equal to the leak volume spill score (LV) divided by the receptors population score (RE), as shown in Eq. 3, Fig. (1) shows the basic risk assessment model.

The relative risk score RRS will be equal to the Index Sum divided by Leak Impact Factor, as shown in Eq. 4. (Muhlbaur, 2004) [18].

$$IS = TPD + C + D + IO \tag{1}$$

$$DF = LV / RE \tag{2}$$

$$LIF = LV \times RE \times DF \times PH \tag{3}$$

$$RRS = IS / LIF \tag{4}$$

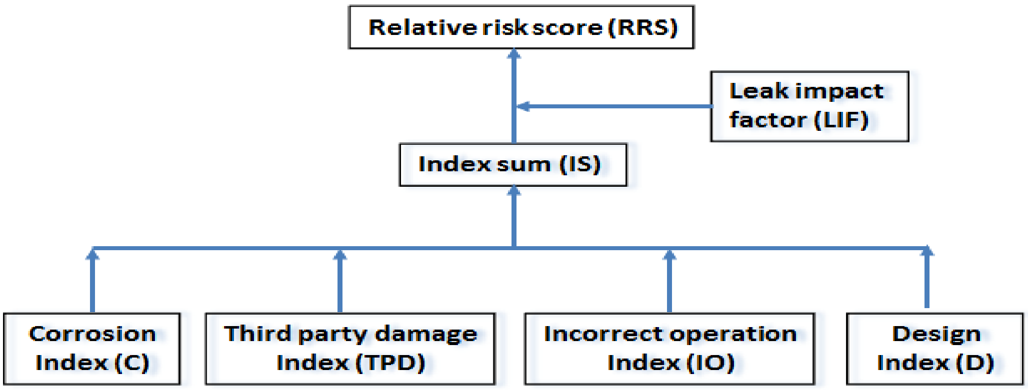


Fig. (1) The basic risk assessment model

FUZZY INFERENCE SYSTEM

Fuzzy set theory

The basic concept of fuzzy set theory was introduced by Zadeh (1965) [4] to work with uncertainty in real world situations. Fuzzy logic is used to deal with problems with unsharp boundaries in which membership is a matter of degree. A fuzzy set defined on a universe of discourse (U) is a characterized by a membership function $\mu(x)$, which takes values from the interval [0, 1] in which 0 points to a non-membership and 1 means a full membership. A membership function provides a measure of the degree of similarity of an element in U to the fuzzy subset. Fuzzy sets are defined for specific linguistic variables. Each linguistic term can be represented by a triangular, trapezoidal or Gaussian shape membership function. The selection of membership function essentially depends on the variable characteristics, available information, and expert's opinion (Wang et al. 2013) [12]. In this work Gaussian membership functions are employed for being the most natural (Markowski & Mannan, 2008) [7], smooth and nonzero at all points (Xie, 2003) [19]. So it can solve complex real world problems with uncertain and vague information such as in risk assessment studies. Gaussian membership function can be represented as shown in Eq. 5.

$$\mu_{A^i}(x) = \exp\left(-\frac{(c_i - x)^2}{2\sigma_i^{\,2}}\right) \tag{5}$$

Where c_i and σ_i are the center and width of the ith fuzzy set A^i, respectively, as shown in Fig. (2).

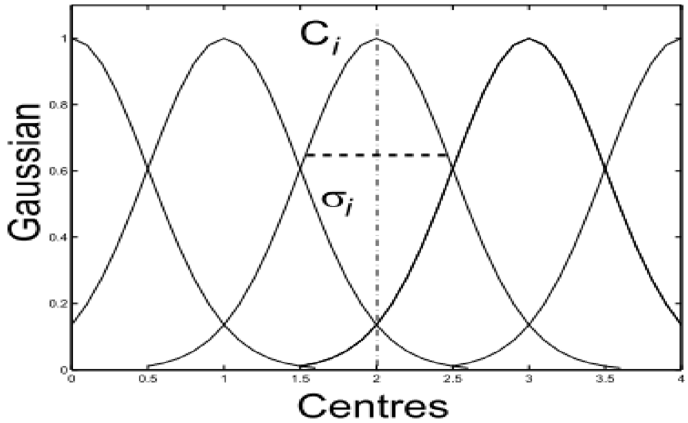


Fig. (2) Gaussian membership functions

Fuzzy inference system is using fuzzy logic to map an input space (universe of discourse) to an output space. The primary mechanism for doing this is a list of IF-THEN rules, membership functions that defines how each point in the input space is mapped to a degree of membership between 0 and 1, and fuzzy logic operators connects with the fuzzy sets. Fuzzy inference system as depicted in Fig. 3. consists of: 1- knowledge base, 2- inference or decision making unit, 3- fuzzification interface, and 4- defuzzification interface.

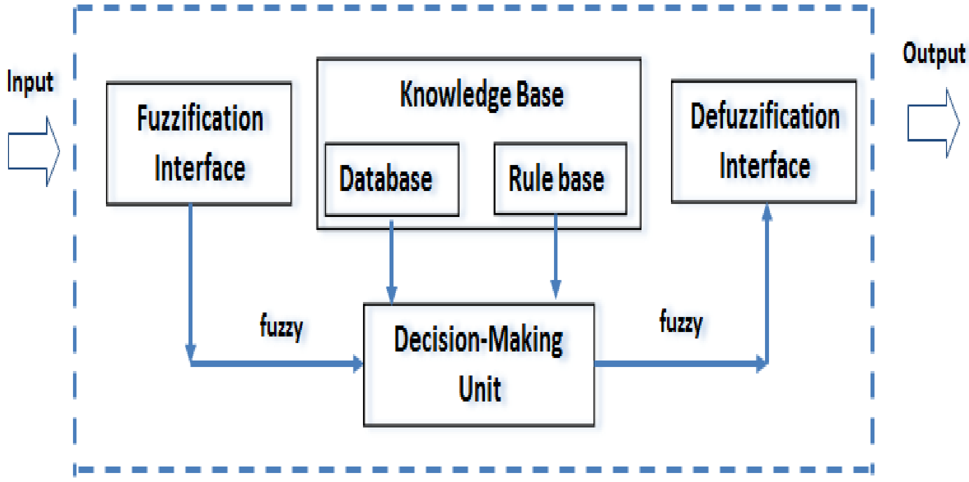


Fig. (3) Fuzzy inference structure

Several fuzzy inference models are employed in many applications, such as Mamdani, Takagi-Sugeno, and Tsukamoto fuzzy model. The Takagi-Sugeno and Mamdani methods are widely used in modeling real world problems. The two methods are quite similar to each other in many aspects. The first two parts of the fuzzy inference process, fuzzification of inputs and applying the fuzzy operators are exactly the same. The main difference is that the Takagi-Sugeno output membership functions are either linear or constant. In this paper, the Takagi-Sugeno method is used for the pipeline risk assessment.

TAKAGI-SUGENO (TS) FUZZY INFERENCE SYSTEM

TS model introduced by (Takagi & Sugeno, 1985) [20] is used to model complex non-linear systems; its main feature is linearization of each fuzzy rule as a linear subsystem. The output is a blend of all these linear subsystems which is done by aggregation of rules.

TS fuzzy model can work with any nonlinear system with a high degree of precision and approved to be a universal approximators of any smooth nonlinear system (Fantuzzi & Rovatti, 1996) [21], (Buckley, 1992) [22].

TS rules use functions of input variables as the rule output (consequent). The general form of TS rule model having two inputs x_1 and x_2, and output U is as follows:

$$\text{if } x_1 \text{ is } A_1 \text{ and } x_2 \text{ is } A_2 \text{ THEN U is } z = f(x_1, x_2)$$

Where $z = f(x_1, x_2)$ is a crisp function of the output; A_1 and A_2 are linguistic terms.

Fig. (4) depicts a typical TS inference mechanism for two input variables.

This function is most commonly linear in which fuzzy rules are linearly generated from a given input-output data, whereas nonlinear function is applied by adaptive techniques (A. Yazdani-Chamzini et. al.,2013) [23].

As mentioned in the previous section of this paper, we have four variables for the Index Sum (IS) model. Which are C, TPD, IO, and D.

The fuzzy IF-THEN rules of this model can be defined as follows:

If (C is ..), AND (TPD is ..), AND (IO is ..), AND (D is ..)
THEN (IS = a×C + b×TPD + c×IO + d×D + e)

And we have four variables for the Leak Impact Factor (LIF) model. Which are PH, DF, LV, and RE.

The fuzzy IF-THEN rules of this model can be defined as follows:

If (PH is ..), AND (LV is ..), AND (RE is ..), AND (DF is ..)
THEN (LIF = f×PH + g×LV + h×RE + i×DF + j)

The parameters a, b, c, d, and e are estimated from the training dataset of the IS model, and the parameters f, g, h, i, and j are estimated from the training dataset of the LIF model.

The final output of the two fuzzy models is the weighted average of all rule outputs in each model, computed as:

$$\text{Final Output} = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i} \tag{6}$$

Where N is the number of rules, w_i is the firing strength to weight the ith fuzzy rule defined as:

$$w_i = \prod_{j=1}^{n} \mu(A_i^j) \tag{7}$$

Where n is the number of input variables; $\mu(A_i^j)$ is the grade of the membership function A_i^j.

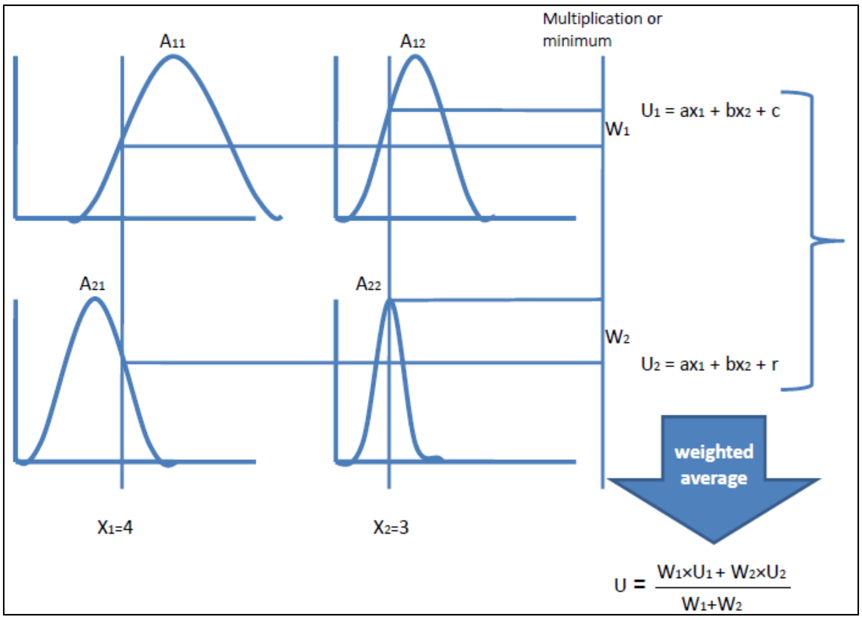


Fig. (4) A typical TS inference mechanism for two input variables

SUBTRACTIVE CLUSTERING

As mentioned before, the risk assessment of pipelines is possible to be modeled qualitatively using the expert's knowledge about the system, and that is done by mathematical modeling from expert's knowledge including the input and output data of the system. Fuzzy clustering technique is a powerful tool for the identification of such system contains possible uncertainty by grouping the input-output data into fuzzy clusters then translating these clusters into fuzzy IF-THEN rules, by this we will avoid identifying all the rules as done in conventional fuzzy inference method. There are different methods for fuzzy clustering; the most common are fuzzy C-means (FCM) clustering (Bezdek, 1981) [24], mountain clustering (Yager & Filev, 1994) [25], and subtractive clustering (Chiu, 1994) [26].

In this work we are using subtractive clustering method. This method has the capability of auto generating the number and the initial location of cluster centers through search techniques, similarly as introduced in mountain clustering method, while fuzzy C-means clustering needs a prior knowledge of the number of clusters. Subtractive clustering has another advantage over mountain clustering as each data point is regarded as a possible cluster center, while mountain clustering regards each grid point as a possible cluster center (Bilgin et al., 2011) [27].

INDEX SUM AND LEAK IMPACT FACTOR DATABASE

Data of the (IS) and the (LIF) models are obtained from experts opinions. Input parameters of (IS) are TPD, C, D, and IO. While input parameters of (LIF) are PH, LV, DF, and RE.

The two models presented in this paper are using a set of statical data consists of 625 input/output data, a part of this data is shown in Table 1 for the (IS) model and Table 2 for the (LIF) model.

PERFORMANCE EVALUATION INDECIS

In order to evaluate the performance of each model, two different indices, including root mean square error (RMSE), and correlation coefficient (R^2); are applied to compare the outputs estimated by the established model with the experts data output. These indices are calculated by the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (A_{\mathrm{i}} - P_i)^2}{N}} \tag{8}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} (A_{\mathrm{i}} - P_i)^2}{\sum_{\mathrm{i}=1}^{\mathrm{N}} (A_{\mathrm{i}} - \overline{A}_i)^2} \tag{9}$$

Where P_i is the predicted values, A_{i} is the qualitative expert's values, $\overline{A}_i$ is the average of the observed set, and N is the number of data set.

The RMSE index, one of the most widely used indices in performance evaluations, is useful to understand the difference between the model output and the actual value. RMSE is a non-negative number that can take zero when the predicted output exactly matches the recorded output and have no upper bound.

R^2 is a positive number that shows how much of the variability in dependent variable can be explained by independent variable(s) and in the other words, how well the model fits the data. R^2 can take values between 0 and 1; which 1 indicates the model can reflect all the variability of the output variable, while 0 expresses that there is a poor correlation between model output and actual output.

Table. (1) Statical description on data set of IS model

No.	*C*	*TPD*	*D*	*IO*	*IS*
1	0	0	0	0	0
2	0	0	0	25	25
3	0	0	0	50	50
4	0	0	0	75	75
5	0	0	0	100	100
6	0	0	25	0	25
7	0	0	25	25	50
8	0	0	25	50	75
9	0	0	25	75	100
10	0	0	25	100	125
11	0	0	50	0	50
12	0	0	50	25	75
13	0	0	50	50	100
14	0	0	50	75	125
15	0	0	50	100	150
16	0	0	75	0	75
17	0	0	75	25	100
18	0	0	75	50	125
19	0	0	75	75	150
20	0	0	75	100	175
21	0	0	100	0	100
22	0	0	100	25	125
23	0	0	100	50	150
24	0	0	100	75	175
25	0	0	100	100	200
⇩					
625	100	100	100	100	400

Table.(2) Statical description on data set of LIF model

No.	*LV*	*RE*	*DF*	*PH*	*LIF*
1	1	1	0.25	0	0
2	1	1	0.25	5.5	1.375
3	1	1	0.25	11	2.75
4	1	1	0.25	16.5	4.125
5	1	1	0.25	22	5.5
6	1	1	1.688	0	0
7	1	1	1.688	5.5	9.284
8	1	1	1.688	11	18.56
9	1	1	1.688	16.5	27.85
10	1	1	1.688	22	37.13
11	1	1	3.125	0	0
12	1	1	3.125	5.5	17.18
13	1	1	3.125	11	34.37
14	1	1	3.125	16.5	51.56
15	1	1	3.125	22	68.75
16	1	1	4.563	0	0
17	1	1	4.563	5.5	25.09
18	1	1	4.563	11	50.19
19	1	1	4.563	16.5	76.28
20	1	1	4.563	22	100.3
21	1	1	6	0	0
22	1	1	6	5.5	33
23	1	1	6	11	66
24	1	1	6	16.5	99
25	1	1	6	22	132
⇩					
625	6	4	6	22	3168

CLUSTERING INDEX SUM AND LEAK IMPACT FACTOR DATA

Against 4 inputs of each model in the index sum and the leak impact factor model there is a single output reflecting the risk decided by the expert knowledge as depicted in Fig. 5, and Fig. 6.

Out of 625 pipeline data for index sum and leak impact factor as shown in Table 1 and Table 2, 500 pipeline data are used for training i.e. to form the membership functions and produce the fuzzy IF-THEN rules, and 125 pipeline data are used for testing and checking the fuzzy model established in each model to validate the model and prevent over fitting that may occur on the training data set.

The pipeline index sum and leak impact factor data are subjected to the subtractive clustering procedure using MATLAB software. The algorithm is repeated for cluster radii 0.1 thru 0.9 in each model. After applying the subtractive clustering method on the training data of the index sum and the leak impact factor with different ranges of cluster radius, the best fitted model in the index sum and the leak impact factor models based on the best performance indices with the testing dataset has a cluster radius 0.8 and 0.6 respectively, as shown in Table 3 and 4. 13 and 27 fuzzy rules are generated in the index sum and leak impact factor models respectively, as depicted in Fig. 7. and Fig. 8. The established index sum model has 65 linear parameters and 104 non linear parameters, while the established leak impact factor model has 295 linear parameters and 472 nonlinear parameters. As we can see from Table 3 and 4, small cluster radius generates many rules and vice versa.

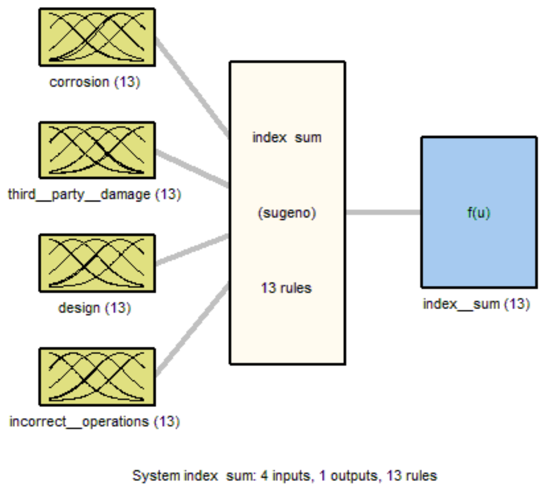


Fig. (5)Index sum fuzzy inference structure

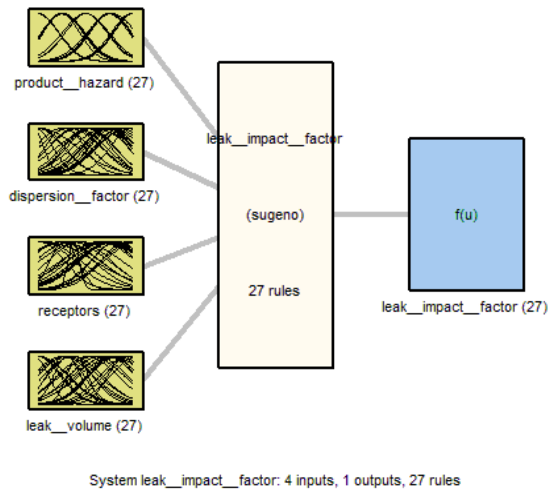


Fig. (6) Leak impact factor fuzzy inference structure

The model performance indices, training RMSE and testing RMSE and the correlation coefficient (R^2) for the best model of index sum (cluster radius = 0.8) obtained are 5.9653e-008 and 7.35411e-008 and 1 respectively. And the model performance indices, training RMSE and testing RMSE and the correlation coefficient (R^2) for the best model of leak impact factor (cluster radius = 0.6) are 1.4065 and 8.7814 and 0.9601 respectively.

The interdependency of input and output parameters derived from the rules generated by subtractive clustering can be shown by using control surfaces as depicted in Fig. 9 and Fig. 10 for index sum and leak impact factor respectively.

As seen in Fig. 9 of index sum model, Fig. 9a shows the interdependency of index sum on design and corrosion, Fig. 9b shows interdependency of index sum on incorrect operations and corrosion, and Fig. 9c shows the interdependency of index sum on third party damage and corrosion. While in Fig. 10 of leak impact factor model, Fig. 10a shows the interdependency of leak impact factor on dispersion factor and product hazard, Fig. 10b shows interdependency of leak impact factor on leak volume and product hazard, and Fig. 10c shows the interdependency of leak impact factor on receptors and product hazard.

Table (3) Index sum's comparative test results of cluster radius value from 0.1 to 0.9 with selected best fitted model

Cluster radius	*Epoch number*	*Number of fuzzy rules*	*Training RMSE*	*Check RMSE*	*Correlation coefficient R^2*
0.1	4	485	0.000597852	5.21032	0.9247
0.2	17	485	0.0016	5.2161	0.9646
0.3	5	379	0.000668717	5.8668	0.9992
0.4	2	127	2.81871e-005	0.583922	1
0.5	12	54	1.4893e-006	2.4321e-006	1
0.6	9	34	3.8899e-007	4.2578e-007	1
0.7	18	20	1.5416e-007	1.5269e-007	1
0.8	185	13	5.9653e-008	7.35411e-008	1
0.9	2	10	5.2631e-008	4.1280e-008	1

Table.(4)Leak impact factor's comparative test results of cluster radius value from 0.1 to 0.9 with selected best fitted model

Cluster radius	*Epoch number*	*Number of fuzzy rules*	*Training RMSE*	*Check RMSE*	*Correlation coefficient R^2*
0.1	1	483	2.5801e-005	66.7306	0.7842
0.2	2	483	7.6480e-005	66.4401	0.9052
0.3	11	339	5.8455e-005	57.9682	0.9421
0.4	82	134	0.000178431	93.7299	0.9368
0.5	78	59	1.15206	27.6552	0.9318
0.6	200	27	1.4065	8.7814	0.9601
0.7	200	17	4.4991	10.5146	0.8734--
0.8	200	12	10.9520	20.8779	-0.9539
0.9	200	10	6.2846	13.0260	-0.9044

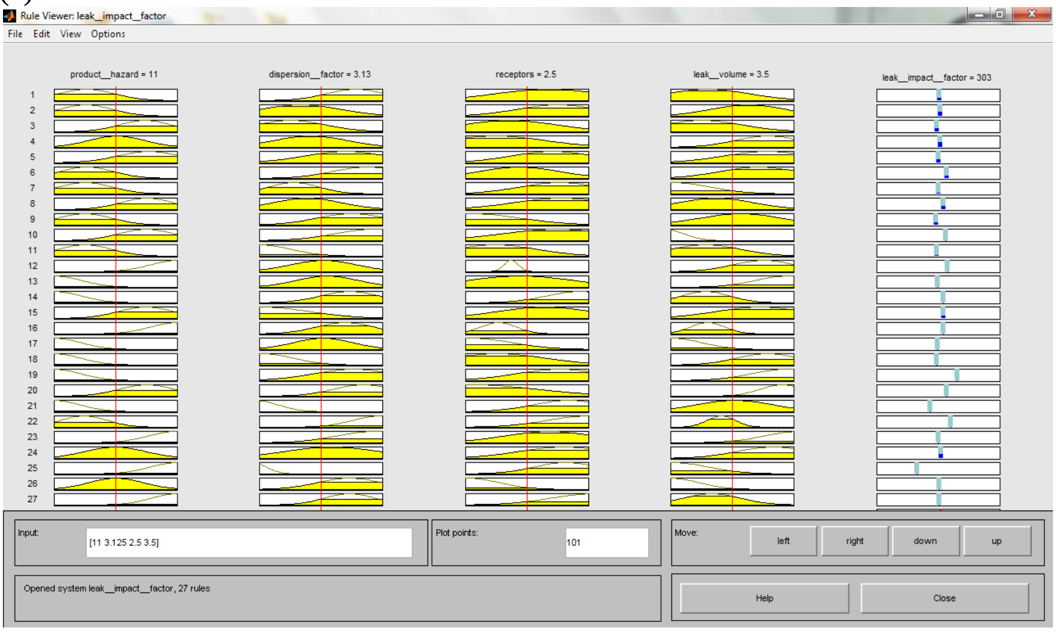
(a)

(b)

1. If (C is in1cluster1) and (TPD is in2cluster1) and (D is in3cluster1) and (IO is in4cluster1) then (IS is out1cluster1) (1)
2. If (C is in1cluster2) and (TPD is in2cluster2) and (D is in3cluster2) and (IO is in4cluster2) then (IS is out1cluster2) (1)
3. If (C is in1cluster3) and (TPD is in2cluster3) and (D is in3cluster3) and (IO is in4cluster3) then (IS is out1cluster3) (1)
4. If (C is in1cluster4) and (TPD is in2cluster4) and (D is in3cluster4) and (IO is in4cluster4) then (IS is out1cluster4) (1)
5. If (C is in1cluster5) and (TPD is in2cluster5) and (D is in3cluster5) and (IO is in4cluster5) then (IS is out1cluster5) (1)
6. If (C is in1cluster6) and (TPD is in2cluster6) and (D is in3cluster6) and (IO is in4cluster6) then (IS is out1cluster6) (1)
7. If (C is in1cluster7) and (TPD is in2cluster7) and (D is in3cluster7) and (IO is in4cluster7) then (IS is out1cluster7) (1)
8. If (C is in1cluster8) and (TPD is in2cluster8) and (D is in3cluster8) and (IO is in4cluster8) then (IS is out1cluster8) (1)
9. If (C is in1cluster9) and (TPD is in2cluster9) and (D is in3cluster9) and (IO is in4cluster9) then (IS is out1cluster9) (1)
10. If (C is in1cluster10) and (TPD is in2cluster10) and (D is in3cluster10) and (IO is in4cluster10) then (IS is out1cluster10) (1)
11. If (C is in1cluster11) and (TPD is in2cluster11) and (D is in3cluster11) and (IO is in4cluster11) then (IS is out1cluster11) (1)
12. If (C is in1cluster12) and (TPD is in2cluster12) and (D is in3cluster12) and (IO is in4cluster12) then (IS is out1cluster12) (1)
13. If (C is in1cluster13) and (TPD is in2cluster13) and (D is in3cluster13) and (IO is in4cluster13) then (IS is out1cluster13) (1)

Fig. (7) (a) rule viewer of the IS fuzzy model and (b) rules generated for the IS fuzzy model

(a)

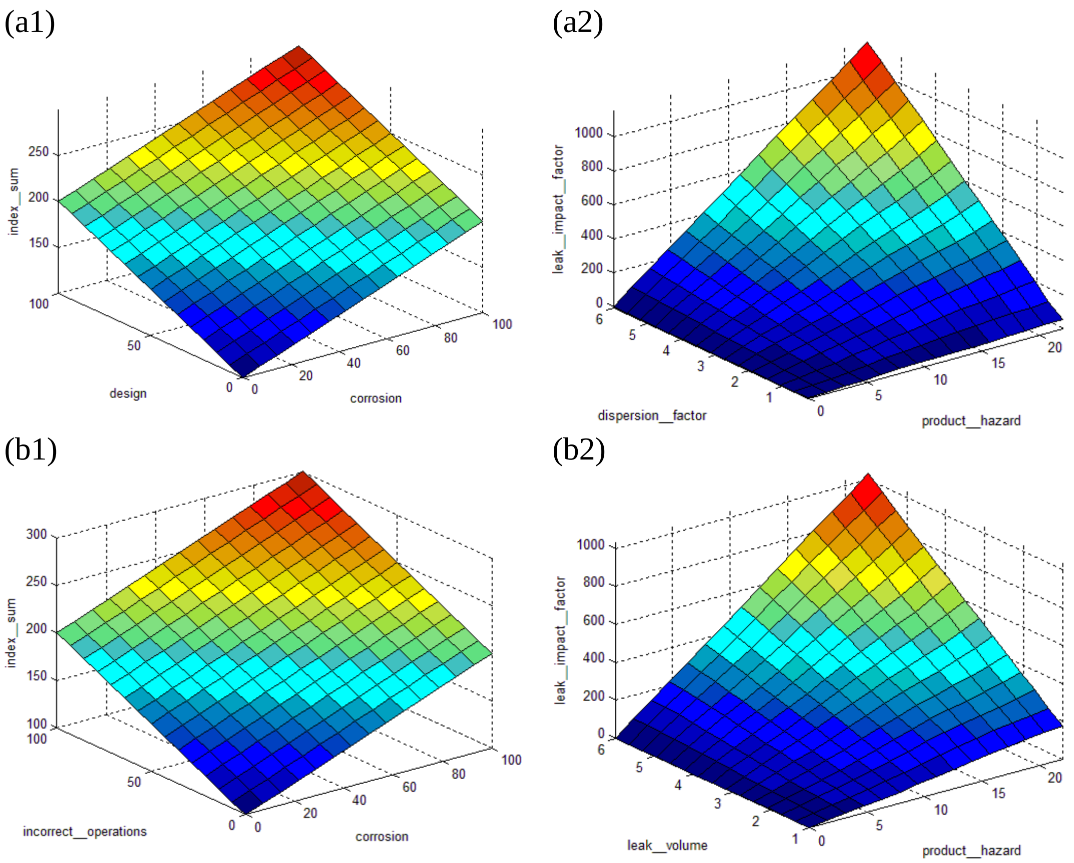


(b)

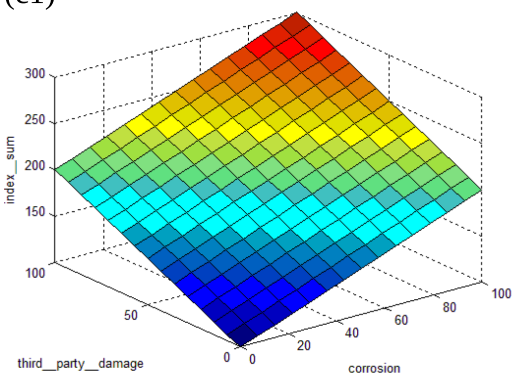
1. If (PH is in1cluster1) and (DF is in2cluster1) and (RE is in3cluster1) and (LV is in4cluster1) then (LIF is out1cluster1) (1)
2. If (PH is in1cluster2) and (DF is in2cluster2) and (RE is in3cluster2) and (LV is in4cluster2) then (LIF is out1cluster2) (1)
3. If (PH is in1cluster3) and (DF is in2cluster3) and (RE is in3cluster3) and (LV is in4cluster3) then (LIF is out1cluster3) (1)
4. If (PH is in1cluster4) and (DF is in2cluster4) and (RE is in3cluster4) and (LV is in4cluster4) then (LIF is out1cluster4) (1)
5. If (PH is in1cluster5) and (DF is in2cluster5) and (RE is in3cluster5) and (LV is in4cluster5) then (LIF is out1cluster5) (1)
6. If (PH is in1cluster6) and (DF is in2cluster6) and (RE is in3cluster6) and (LV is in4cluster6) then (LIF is out1cluster6) (1)
7. If (PH is in1cluster7) and (DF is in2cluster7) and (RE is in3cluster7) and (LV is in4cluster7) then (LIF is out1cluster7) (1)
8. If (PH is in1cluster8) and (DF is in2cluster8) and (RE is in3cluster8) and (LV is in4cluster8) then (LIF is out1cluster8) (1)
9. If (PH is in1cluster9) and (DF is in2cluster9) and (RE is in3cluster9) and (LV is in4cluster9) then (LIF is out1cluster9) (1)
10. If (PH is in1cluster10) and (DF is in2cluster10) and (RE is in3cluster10) and (LV is in4cluster10) then (LIF is out1cluster10) (1)
11. If (PH is in1cluster11) and (DF is in2cluster11) and (RE is in3cluster11) and (LV is in4cluster11) then (LIF is out1cluster11) (1)
12. If (PH is in1cluster12) and (DF is in2cluster12) and (RE is in3cluster12) and (LV is in4cluster12) then (LIF is out1cluster12) (1)
13. If (PH is in1cluster13) and (DF is in2cluster13) and (RE is in3cluster13) and (LV is in4cluster13) then (LIF is out1cluster13) (1)
14. If (PH is in1cluster14) and (DF is in2cluster14) and (RE is in3cluster14) and (LV is in4cluster14) then (LIF is out1cluster14) (1)
15. If (PH is in1cluster15) and (DF is in2cluster15) and (RE is in3cluster15) and (LV is in4cluster15) then (LIF is out1cluster15) (1)
16. If (PH is in1cluster16) and (DF is in2cluster16) and (RE is in3cluster16) and (LV is in4cluster16) then (LIF is out1cluster16) (1)

17. If (PH is in1cluster17) and (DF is in2cluster17) and (RE is in3cluster17) and (LV is in4cluster17) then (LIF is out1cluster17) (1)
18. If (PH is in1cluster18) and (DF is in2cluster18) and (RE is in3cluster18) and (LV is in4cluster18) then (LIF is out1cluster18) (1)
19. If (PH is in1cluster19) and (DF is in2cluster19) and (RE is in3cluster19) and (LV is in4cluster19) then (LIF is out1cluster19) (1)
20. If (PH is in1cluster20) and (DF is in2cluster20) and (RE is in3cluster20) and (LV is in4cluster20) then (LIF is out1cluster20) (1)
21. If (PH is in1cluster21) and (DF is in2cluster21) and (RE is in3cluster21) and (LV is in4cluster21) then (LIF is out1cluster21) (1)
22. If (PH is in1cluster22) and (DF is in2cluster22) and (RE is in3cluster22) and (LV is in4cluster22) then (LIF is out1cluster22) (1)
23. If (PH is in1cluster23) and (DF is in2cluster23) and (RE is in3cluster23) and (LV is in4cluster23) then (LIF is out1cluster23) (1)
24. If (PH is in1cluster24) and (DF is in2cluster24) and (RE is in3cluster24) and (LV is in4cluster24) then (LIF is out1cluster24) (1)
25. If (PH is in1cluster25) and (DF is in2cluster25) and (RE is in3cluster25) and (LV is in4cluster25) then (LIF is out1cluster25) (1)
26. If (PH is in1cluster26) and (DF is in2cluster26) and (RE is in3cluster26) and (LV is in4cluster26) then (LIF is out1cluster26) (1)
27. If (PH is in1cluster27) and (DF is in2cluster27) and (RE is in3cluster27) and (LV is in4cluster27) then (LIF is out1cluster27) (1)

Fig. (8) (a) rule viewer of the LIF fuzzy model and (b) rules generated for the LIF fuzzy model



(c1)



(c2)

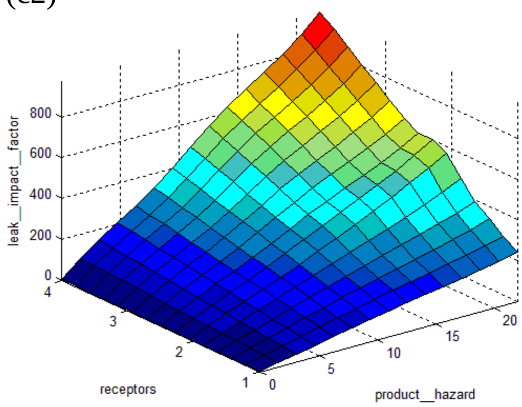


Fig. (9) Control surface of index sum on (a1) design and corrosion; (b1) incorrect operations and corrosion; (c1) third party damage and corrosion

Fig. (10) Control surface of leak impact factor on (a2) dispersion factor and product hazard; (b2) leak volume and product hazard; (c2) receptors and product hazard

CASE STUDY

A typical case study is used to demonstrate the proposed approach for pipeline risk assessment is presented. The information is taken from interviews with pipeline operators of SUMED pipeline located in Egypt (Fig. 11). The SUMED pipeline is crucial for the international energy market as it allows the transport of exported crude oil, transported by very large crude carriers VLCCs, coming from Gulf countries and passing through the Suez Canal, on their way to Europe and/or USA. These large tankers cannot pass through the Suez Canal in fully loaded condition as their draft will exceed the depth of the Canal. Prior to passing the Canal, the loaded tankers are moored to a single point mooring system (SPM) at the Ain Sukhna terminal. Crude oil is then discharged from the tanker to the pipeline via the SPM piping system. The tankers can then pass the Canal in ballast condition with a low draft. Crude oil is dispatched through two parallel pipelines, 42 inches diameter 320 km long, starting from Ain Sukhna terminal to Sidi Kerir terminal crossing the river Nile South of Cairo where a pressure relief station preserves the pipeline from any over pressure. An intermediate boosting station, consisting of six gas turbine driven pumps, located midway at Dahshour is used to help push the oil throughout to its final destination, the Sidi Kerir terminal. After passing the canal in ballast condition, the tankers are moored to a single point mooring system (SPM) at the Sidi Kerir terminal where the oil is reloaded to the tanker via the terminal pumps and through the SPM piping system.

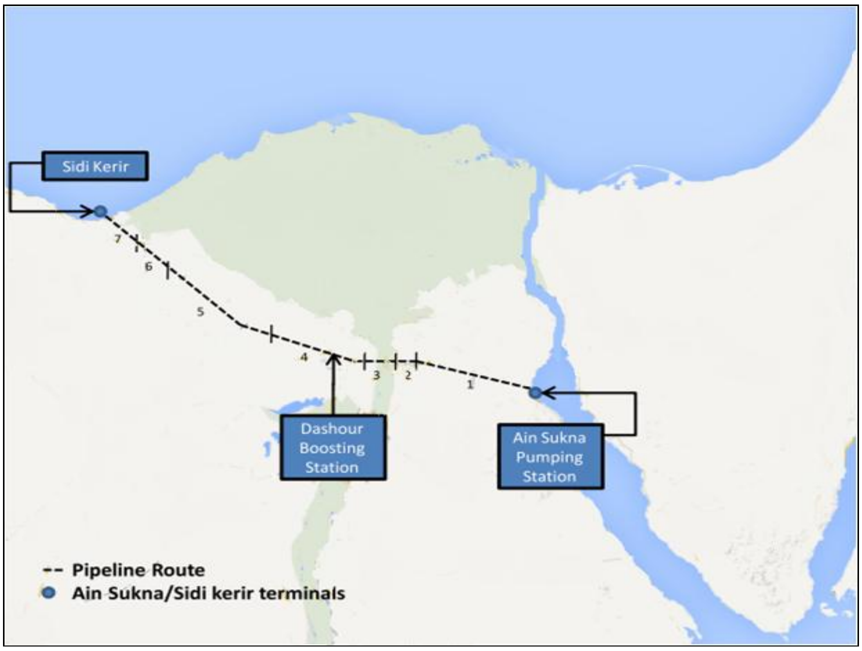


Fig. (11) SUMED pipeline sections

The pipeline has different wall thicknesses varied from 11.13 mm to 22.22 mm according to design. The 320 km pipeline will be sectioned to 7 sections with varied distances. Sectioning the pipeline is done by putting the following phenomenon in consideration; the type of land, soil condition, atmospheric type, population density, Crossing River & water ways, high/low lands, and finally the existence of Right of Way (ROW). Sections will be with the following distances & characteristics as presented in Table 5.

Table. (5) Pipeline sections

Section number	*Starts-Ends*	*Pipeline length*	*Characteristics*
1	0 km – 100 km	100 km	Starts from Ain Sukna, lowest point on land, desert area.
2	100 km – 105 km	5 km	Passing near a cement factory, high population density.
3	105 km – 115 km	10 km	Passing through the river Nile, no ROW, presence of seasonal corps, high population density.
4	115km –165km	50 km	Moderate population density.
5	165 km – 265 km	100 km	Low population density.
6	265 km – 295 km	30 km	Presence of seasonal corps, no ROW.
7	295 km – 320 km	25 km	Ends at Sedi Kerir, passing through lake, high population density.

The risk assessment is performed by using the traditional method and by the proposed model on each pipeline section separately, the results of both methods are compared in the next section of the paper. The section pipeline with the lowest RRS value is selected as the riskiest section, that may help the pipeline operators to start managing the risk on the lowest score pipeline section. In the lowest RRS value section the operator may start with the lowest scored index to improve the reliability and safety of this section, e.g. low scored design index. The results of the traditional RRS method for risk assessment of 7 sections are calculated, based on Eq. 1, 2, 3, and 4. An example is presented as follows:

$IS_{(section1)} = 84 + 83 + 1 + 82 = 250$
$DF_{(section1)} = 2/2 = 1$
$LIF_{(section1)} = 9\times2\times1\times2 = 36$
$RRS_{(section1)} = \frac{IS_{(section1)}}{LIF_{(section1)}} = 250/36 = 6.94$

RESULTS AND DISCUSSIONS

The output relative risk score RRS results of the proposed model (index sum, and leak impact factor) are presented in Table 6. Including the entry values of index sum; third party damage, corrosion, design, and incorrect operations. And the entry values of leak impact factor; product hazard, leak volume, dispersion factor, and receptors.

Table. (6) Output RRS results of the proposed model

Section No.	*TPD*	*C*	*D*	*IO*	*PH*	*LV*	*DI*	*RE*	*IS*	*LIF*	*RRS*	*Rank*
1	84	83	1	82	9	2	1	2	250	33.7	7.418	4
2	77	81	30	82	9	2	0.6	3	270	35.2	7.670	5
3	68	72.5	32	87	9	3	0.75	4	260	84.9	3.062	1
4	77	84	36.5	87	9	2	0.6	3	285	35.2	8.096	6
5	84	83	36.5	82	9	2	1	2	286	33.7	8.486	7
6	64	81.5	36.5	82	9	3	1.5	2	264	76.8	3.437	3
7	63	70	37	84	9	3	1.5	2	254	76.8	3.307	2

As we notice from Table 6. Section number 3 of the pipeline has the lowest RRS value and ranked as the riskiest part of the pipeline as it passes through the river Nile. Section number 3 will be the starting point in risk management to decrease the risks on it. The risk assessor can start by enhancing the design index record of this section as it has the lowest value between the index sum indices.

The design index record can be enhanced by doing the following:

• Increase Pipe Safety Factor.
• Increase System Safety Factor.
• Avoid Fatigue.

• Avoid Surge Potential.
• Make a System Hydrotest to ensure pipeline integrity.
• Avoid Pipe Movements.

To compare the output RRS results of the proposed model with those of the traditional method; the output values of RRS and its ranks in both methods are presented in Table 7. The correlation between the traditional method output RRS values and the proposed model output RRS values is depicted in Fig. 12. The results demonstrate that the proposed fuzzy model by subtractive clustering is a powerful tool for pipeline risk assessment.

Table. (7)Output RRS results of traditional method and proposed model

Section	*Traditional method*						*Proposed model*					
number	*IS*	*Rank*	*LIF*	*Rank*	*RRS*	*Rank*	*IS*	*Rank*	*LIF*	*Rank*	*RRS*	*Rank*
1	250	7	36	2	6.94	4	250`	7	33.7	4	7.418	4
2	270	3	32.4	3	8.3	6	270	3	35.2	3	7.670	5
3	259.5	5	81	1	3.2	2	260	5	84.9	1	3.062	1
4	284.5	2	32.4	3	8.78	7	285	2	35.2	3	8.096	6
5	286	1	36	2	7.93	5	286	1	33.7	4	8.486	7
6	264	4	81	1	3.259	3	264	4	76.8	2	3.437	3
7	254	6	81	1	3.1358	1	254	6	76.8	2	3.307	2

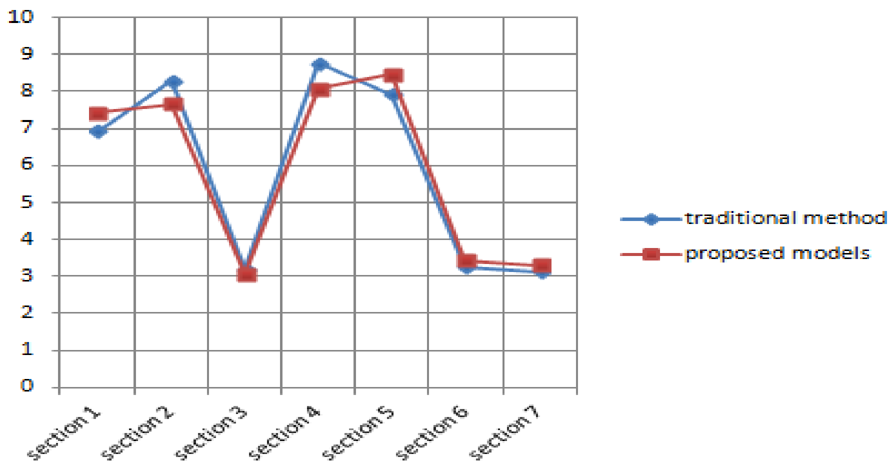


Fig. (12) RRS results of traditional method versus proposed models

In order to further study the relationship between the qualitative method and the proposed model, the degree of correlation of the index sum and leak impact factor obtained by the proposed subtractive clustering fuzzy model with those of the qualitative method was calculated as:

$$\rho = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{10}$$

Where:

x = qualitative output results
y = fuzzy inference output results
$\bar{x}$ = mean value of x
$\bar{y}$ = mean value of y
$\mathrm{cov}(x, y)$ = covariance of x and y
σ_x = standard deviation of x
σ_y = standard deviation of y

As depicted in Fig. (13) The high correlation coefficient value (ρ = 0.9999) for index sum, (ρ = 0.9903) for leak impact factor, and (ρ = 0.9821) for RRS, implies the effectiveness of using the TS fuzzy inference method based on subtractive clustering.

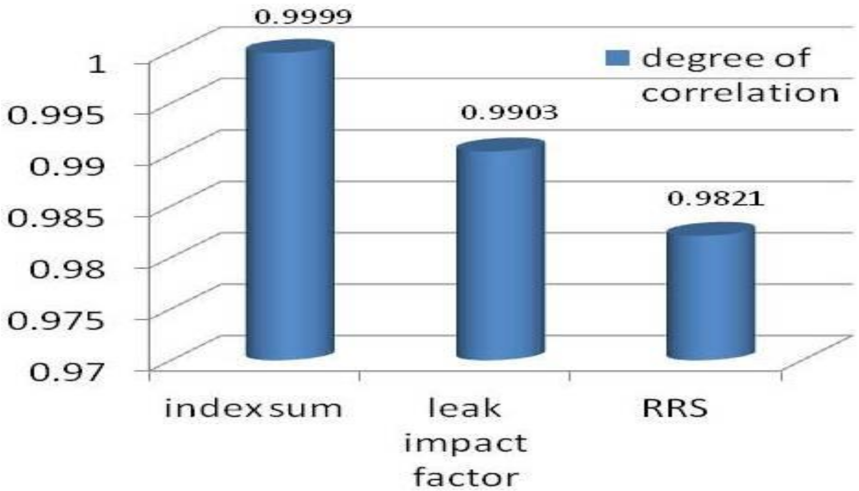


Fig. (13) Correlation coefficient degree between qualitative and subtractive clustering method

CONCLUSION

In this paper the indexing pipeline risk assessment methodology is integrated with subtractive clustering fuzzy logic to deal with the uncertainty of the real world conditions and to avoid the difficulties of constructing many rules as the computational complexity increases with the dimensions of the system variables because the number of rules increases exponentially as the number of system variables increases.

A case study of a petroleum pipeline is used to demonstrate the proposed approach for pipeline risk assessment where the results of the proposed model are compared with the qualitative method. The pipeline located in Egypt is transporting crude oil from Ain Sukhna terminal to Sidi Kerir terminal. The pipeline is divided to seven sections and the risk assessment procedure is done for each section by both qualitative and proposed model. The results showed that the computed RRS values using proposed model are consistent with those obtained using qualitative method. The results also showed a high correlation and high accuracy of the proposed model. The proposed model is evaluated using training RMSE, testing RMSE, and R^2 of values 5.9653e-008 and 7.35411e-008 and 1 for index sum model, and 1.4065 and 8.7814 and 0.9601 for the leak impact factor model respectively.

REFERENCES

1. Washington, DC – February 7, 2012 – The Association of Oil Pipe Lines (AOPL) (aopl report)

2. Shahriar, A., Sadiq, R., & Tesfamariam,S. (2012). Risk analysis for oil & gas pipelines: a sustainability assessment approach using fuzzy based bow-tie analysis. J. of Loss Prevention in the Process Industries, 25, 505-523.
3. Transmission Pipelines and Land Use: A Risk-Informed Approach -- Special Report 281. Washington, DC: The National Academies Press, 2004.
4. Zadeh, L. A. (1965). Fuzzy set. Inf. and Control, 8, 338-353.
5. Elsayed, T. (2009). Fuzzy inference system for the risk assessment of liquefied natural gas carriers during loading/offloading at terminals. Appl. Ocean Res., 31, 179-185.
6. M.-L. Lee, H.-Y. Chung, F.-M. Yu. (2003). Modeling of hierarchical fuzzy systems. Fuzzy Sets and Syst., 138, 343-361.
7. Markowski, A. S., & Mannan, M. S. (2008). Fuzzy risk matrix. J. of Hazard. Mater., 159, 152-157.
8. Markowski, A. S., & Mannan, M. S. (2009). Fuzzy logic for piping risk assessment (pfLOPA). J. of Loss Prevention in the Process Industries, 22, 921-927.
9. Jamshidi, A., Yazdani-Chamzini, A., Haji Yakhchali, S. & Khaleghi, S. (2013). Developing a new fuzzy inference system for pipeline risk assessment. J. of Loss Prevention in the Process Industries, 26, 197-208.
10. Chandima Ratnayake, R.M. (2014) . Application of a fuzzy inference system for functional failure risk rank estimation: RBM of rotating equipment and instrumentation. J. of Loss Prevention in the Process Industries 26, 216-224.
11. Sa'idi, E., Anvaripour, B., Jaderi, F., Nabhani, N. (2014). Fuzzy Risk Modeling of Process Operations in the Oil and Gas Refineries, J. of Loss Prevention in the Process Industries, 30, 63-73.
12. Wang, D., Zhang, P., Chen, L., (2013). Fuzzy fault tree analysis for fire and explosion of crude oil tanks. J. of Loss Prevention in the Process Industries 26, 1390-1398.
13. Khalil, M., Abdou, M.A., Mansour, M.S., Farag H.A., Ossman, M.E., (2012). A cascaded fuzzy-LOPA risk assessment model applied in natural gas industry. J. of Loss Prevention in the Process Industries 25, 877-882.
14. Markowski, A.S., Mannan, M.S., Bigoszewska, A. (2009). Fuzzy logic for process safety analysis. J. of Loss Prevention in the Process Industries 22, 695-702.
15. Yuhua, D., Datao, Y. (2005). Estimation of failure probability of oil and gas transmission pipelines by fuzzy fault tree analysis. J. of Loss Prevention in the Process Industries 18, 83-88.
16. Aqlan, F., Mustafa Ali, E. (2014). Integrating lean principles and fuzzy bow-tie analysis for risk assessment in chemical industry. J. of Loss Prevention in the Process Industries 29, 39-48.
17. Amindoust, A., Ahmed, S., Saghafiniab, A., Bahreininejad, A. (2012). Sustainable supplier selection: A ranking model based on fuzzy inference system. Appl. Soft Computing 12, 1668-1677.
18. Muhlbauer, W. K. (2004). Pipeline risk management manual; ideas, techniques, and resources (1st ed.). Elsevier Inc.

19. [19] Xie, M. (2003). Fundamentals of robotics: Linking perception to action. London: World Scientific Publishing Co Ltd.
20. [20] Takagi, T., Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. IEEE Trans. on Syst., Man, and Cybern. 15 (1), 116 – 132.
21. C. Fantuzzi, C., Rovatti, R., (1996). On the approximation capabilities of the homogeneous Takagi-Sugeno model, Fuzzy Systems, Proceedings of the Fifth IEEE International Conference, 2.
22. Buckley, J., (1992). Universal fuzzy controllers, Automatica (J. of IFAC), 28, (6), 1245-1248.
23. Yazdani-Chamzini, A., Razani, M., Haji Yakhchali, S. (2013). Developing a fuzzy model based on subtractive clustering for road header performance prediction. Automation in Construction 35, 111-120.
24. Bezdec, J.C., (1981). Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York.
25. Yager, R.R., Filev,D.P. (1994). Approximate clustering via the mountain method, IEEE Trans. Syst., Man, Cybern., 24, (8), 1279-1284.
26. Chiu, S.L., (1994). Fuzzy Model Identification Based on Cluster Estimation, J. of Intell. & Fuzzy Syst., 2, (3), 267–278.
27. Bilgin, G., Ertürk, S., Tülay Yildirim, T. (2011). Segmentation of Hyperspectral Images via Subtractive Clustering and Cluster Validation Using One-Class Support Vector Machines. IEEE Trans. on Geoscience and Remote Sens., 49, (8), 2936-2944.