Presentation Title: Risk Assessment of Cross Country Pipelines using Fuzzy Clustering.

Name: Essam Mohamed Abdel Salam.

Outline:

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.

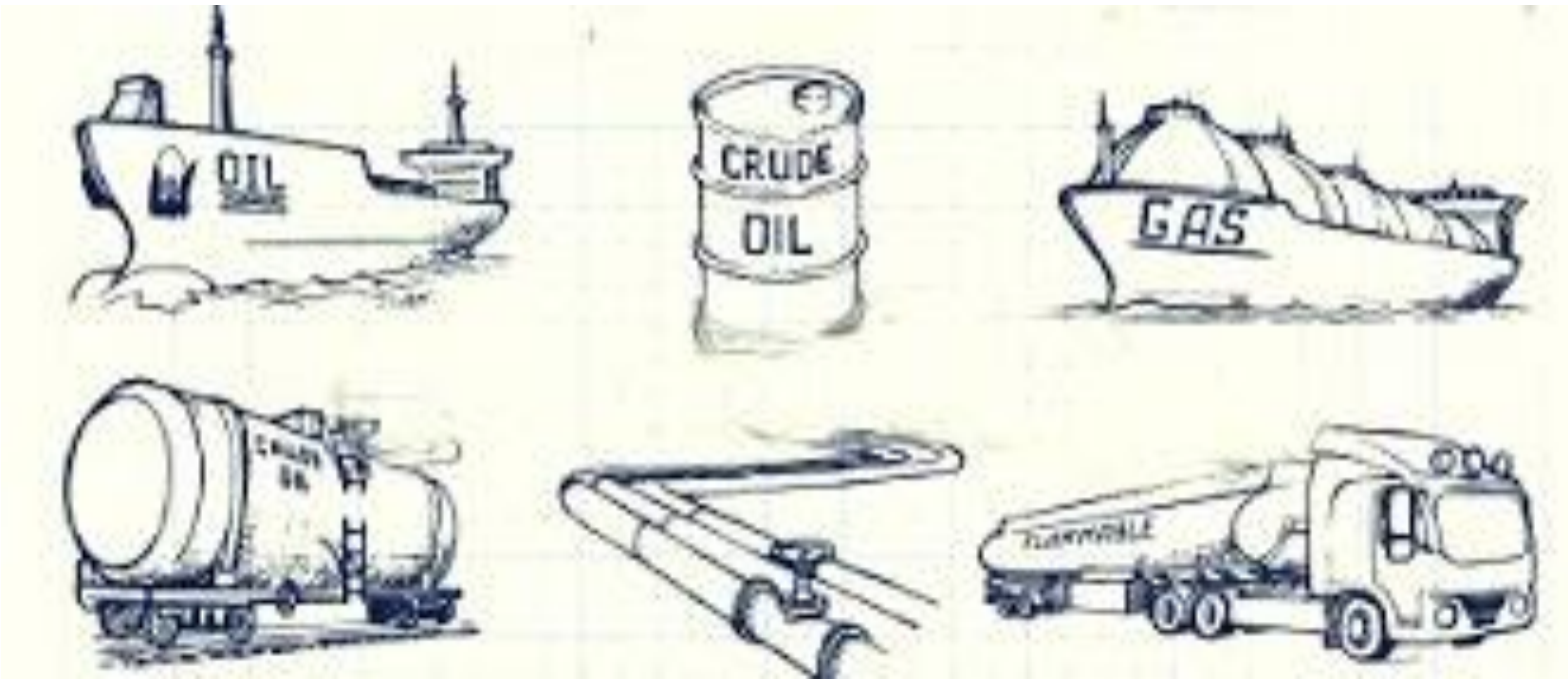


Fig.1. Means of petroleum transportation

Pipelines move huge quantities of crude oil and refined products efficiently, dependably and at a low cost, making many petroleum-based products more affordable for Egypt consumers.

In addition, by reducing the volume of crude oil and refined products that must be moved by truck or railway, pipelines make the roads and railways safer for travelers.

- 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.

Fig.2. Pipeline leakage in Alberta Canada

Company not even aware of the leak until locals called to complain. Red Deer, Alberta, Canada river watershed polluted.

- 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.
- 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.

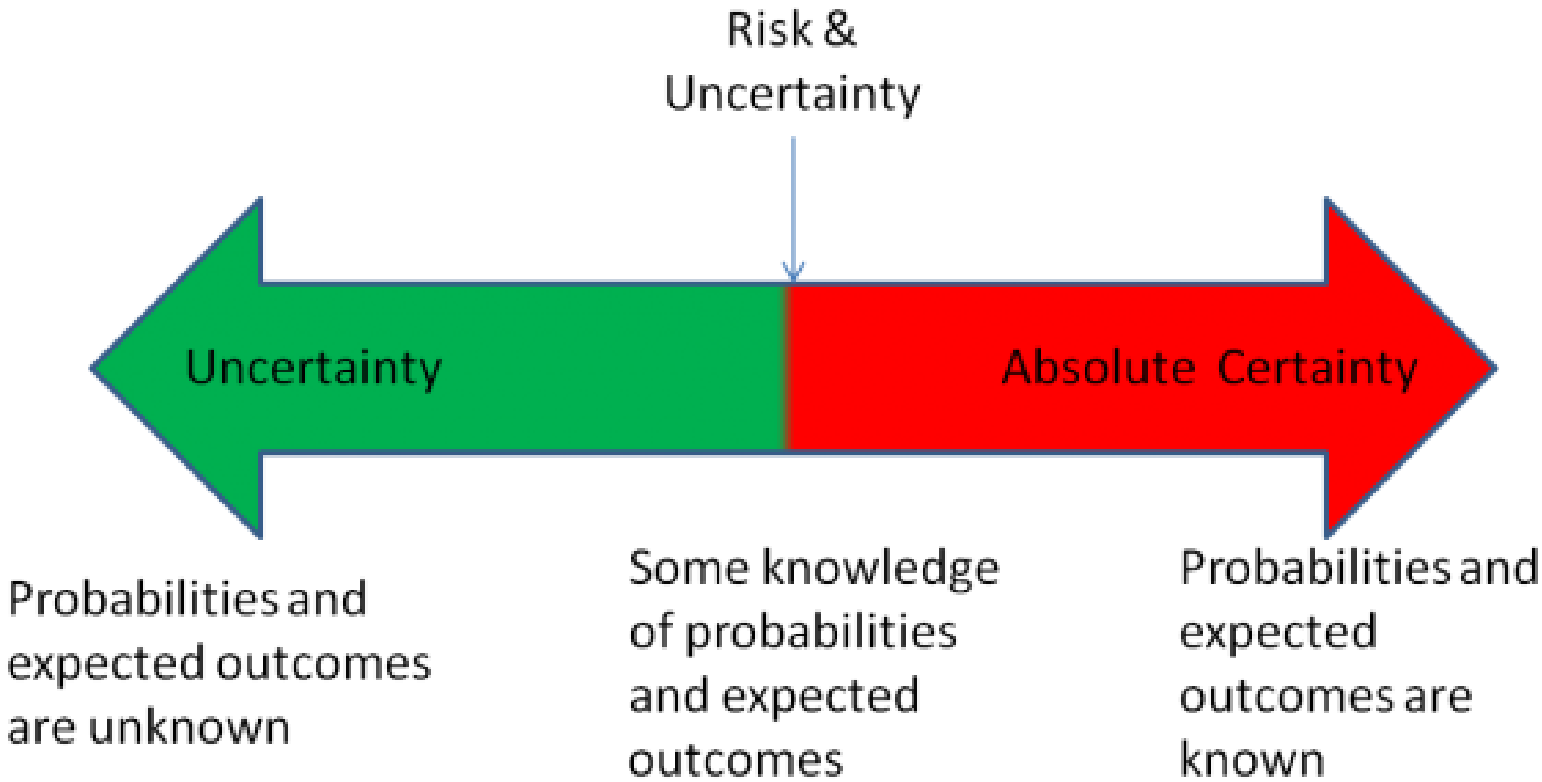


Fig.3. Risk assessment and the uncertainty problem.

- 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) 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.

- 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 RMSE, Check RMSE, and correlation coefficient (R^2).

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.

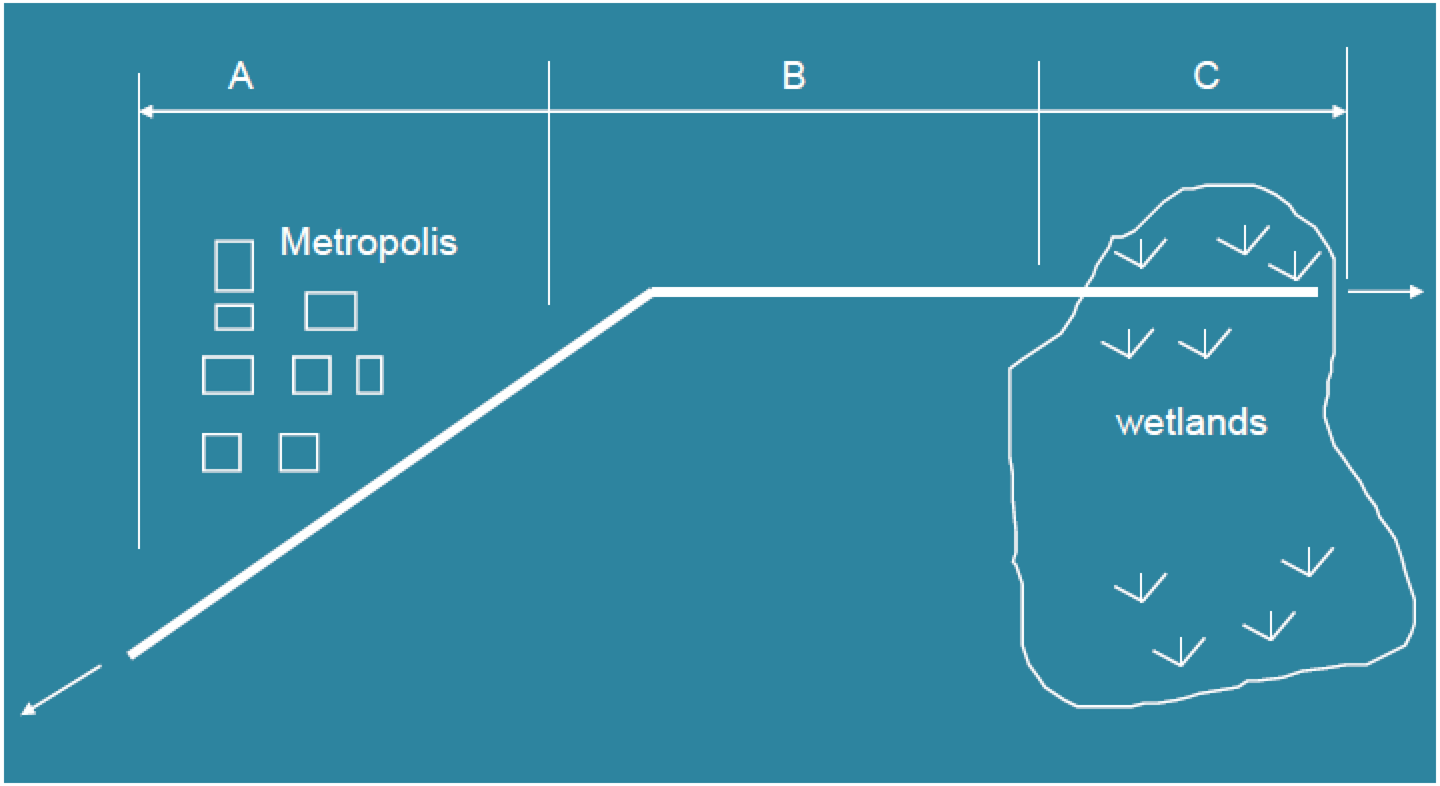
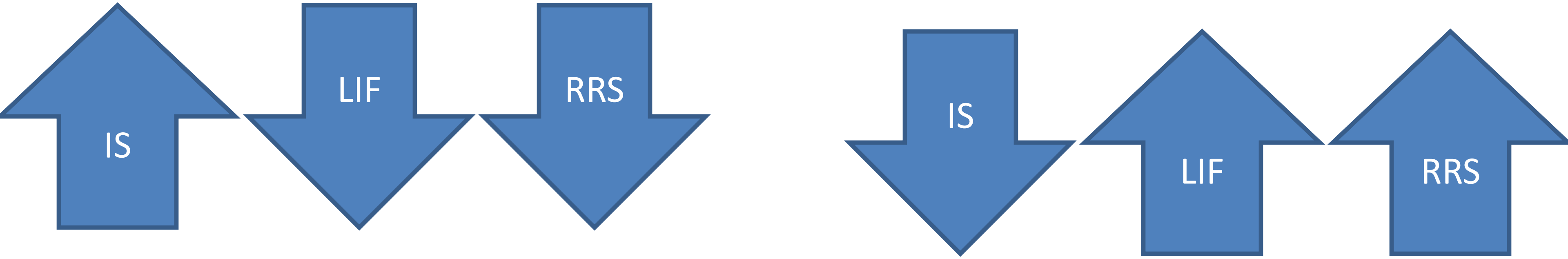


Fig.4. Pipeline sectioning.

- 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.



- $IS = TPD + C + D + IO \quad (1)$
- $DF = LV / RE \quad (2)$
- $LIF = LV \times RE \times DF \times PH \quad (3)$
- $RRS = IS / LIF \quad (4)$

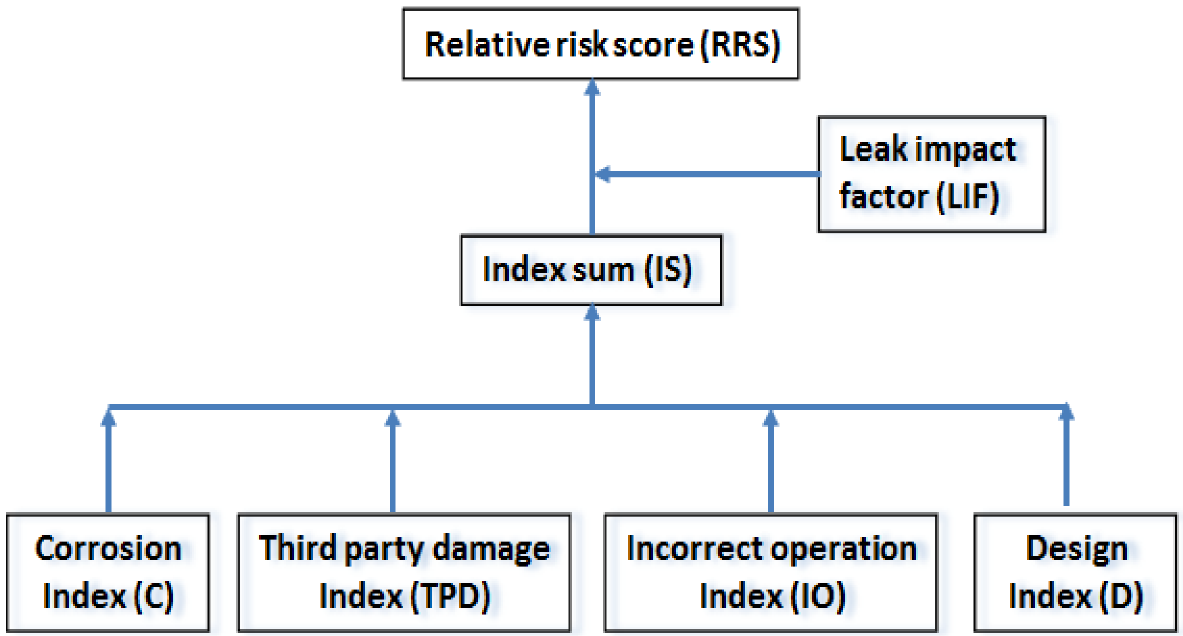


Fig.5. The basic risk assessment model

3. Fuzzy inference system

- A Fuzzy Inference System (FIS) is a way of mapping an input space to an output space using fuzzy logic.
- FIS uses a collection of fuzzy membership functions and rules.
- The rules in FIS (*IF-THEN*) are in the form:
  - *if p then q,* where p and q are fuzzy statements.

  For example, in a fuzzy rule
  - if x is low and y is high then z is medium.
  - Here x is low; y is high; z is medium are linguistic terms; x and y are input variables; z is an output variable, low, high, and medium are fuzzy sets.

- The functional operations in fuzzy inference system proceed in the following steps.
  - Fuzzification
  - Fuzzy Inferencing (apply implication method)
  - Aggregation of all outputs
  - Defuzzification

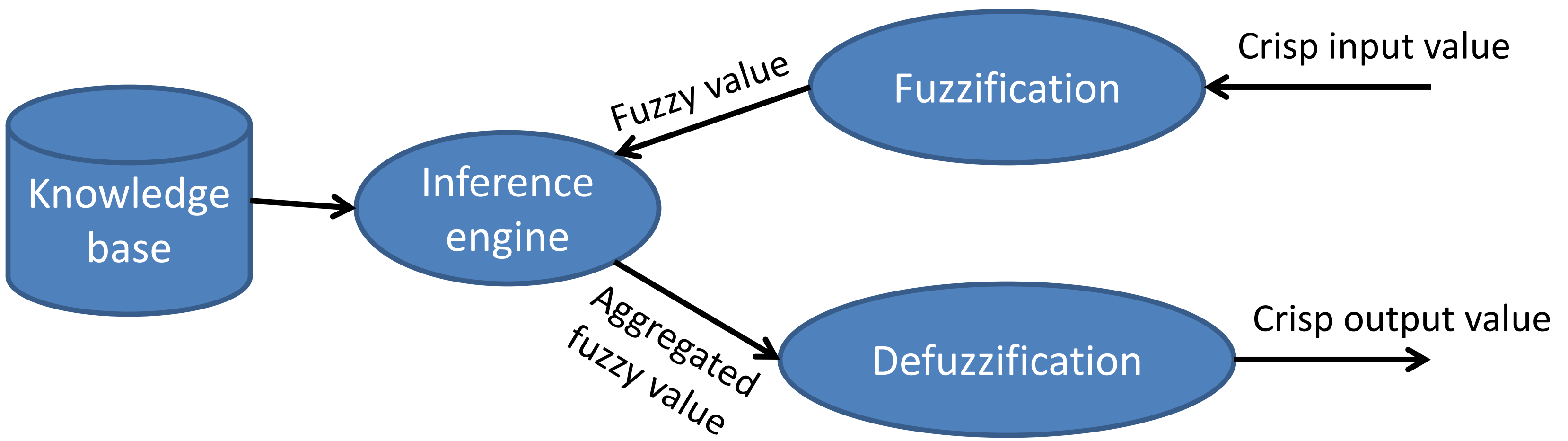


Fig.6. Fuzzy inference structure

- TS model introduced by (Takagi & Sugeno, 1985) 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 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:

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

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

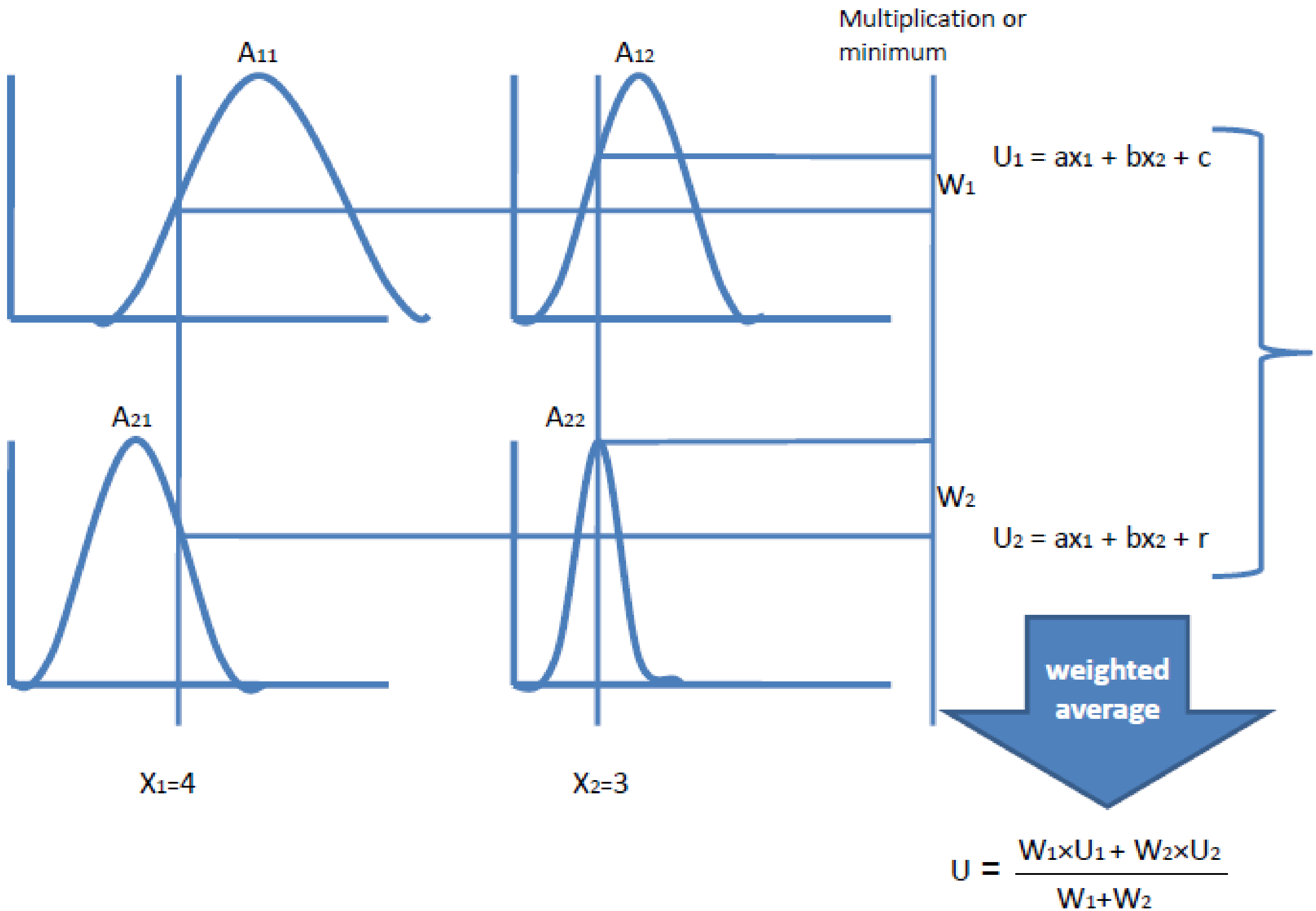


Fig.7. A typical inference mechanism for two input variables

- 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:

$$if\ (C\ is\ ..), AND\ (TPD\ is\ ..), AND (IO\ is\ ..), AND\ (D\ is\ ..)$$

$$THEN\ (IS = a \times C + b \times TPD + c\ \times IO + d \times D + e)$$

The parameters a, b, c, d, and e are estimated from the training dataset of the IS model.

- 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:

$$if\ (PH\ is\ ..),AND\ (LV\ is\ ..),AND(RE\ is\ ..),AND\ (DF\ is\ ..)$$
$$THEN\ (LIF = f \times PH + g \times LV + h\ \times RE + i \times DF + j)$$

  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:
- $\text{Final Output} = \frac{\sum_{i=1}^{N} w_i z_i}{\sum_{i=1}^{N} w_i}$ (6)

  Where N is the number of rules, w_i is the firing strength to weight the i^{th} fuzzy rule defined as:
- $w_i = \prod_{j=1}^{n} \mu\,(A_i^j)$ (7)

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

4. Subtractive clustering

- Pipeline risk assessment by mathematical modeling from expert's knowledge including the input and output data of the system is possible.
- 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.

5. 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.

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.1. Statical description on dataset of IS 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	18.568
9	1	1	1.688	16.5	27.852
10	1	1	1.688	22	37.136
11	1	1	3.125	0	0
12	1	1	3.125	5.5	17.1875
13	1	1	3.125	11	34.375
14	1	1	3.125	16.5	51.5625
15	1	1	3.125	22	68.75
16	1	1	4.563	0	0
17	1	1	4.563	5.5	25.0965
18	1	1	4.563	11	50.193
19	1	1	4.563	16.5	75.2895
20	1	1	4.563	22	100.386
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

Table.2. Statical description on dataset of LIF model

6. Performance evaluation indices

- $$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(A_i - P_i)^2}{N}} \quad (8)$$
- $$R^2 = 1 - \frac{\sum_{i=1}^{N}(A_i - P_i)^2}{\sum_{i=1}^{N}(A_i - \bar{A}_i)^2} \quad (9)$$
- Where P_i is the predicted values, A_i is the the qualitative expert's values, $\bar{A}_i$ is the average of the observed set, and N is the number of data set.
- RMSE equals zero when the predicted output exactly matches the recorded output.
- If R^2 equals zero it reflects a very bad correlation, if equals one it reflects 100% correlation.

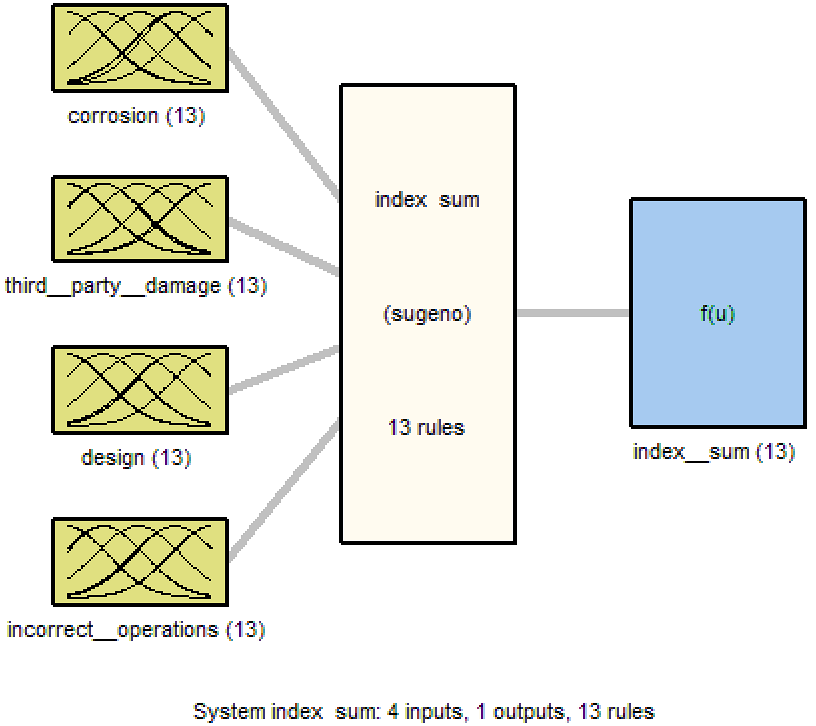
7. Clustering the pipeline index sum and leak impact factor data

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.3. Best selected model of IS

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

Table.4. Best selected model of LIF



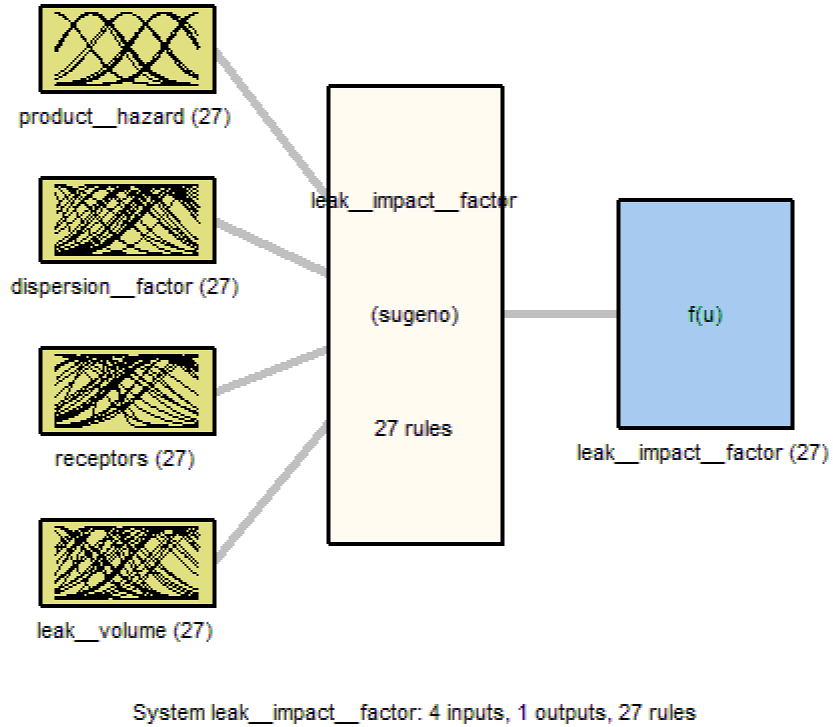
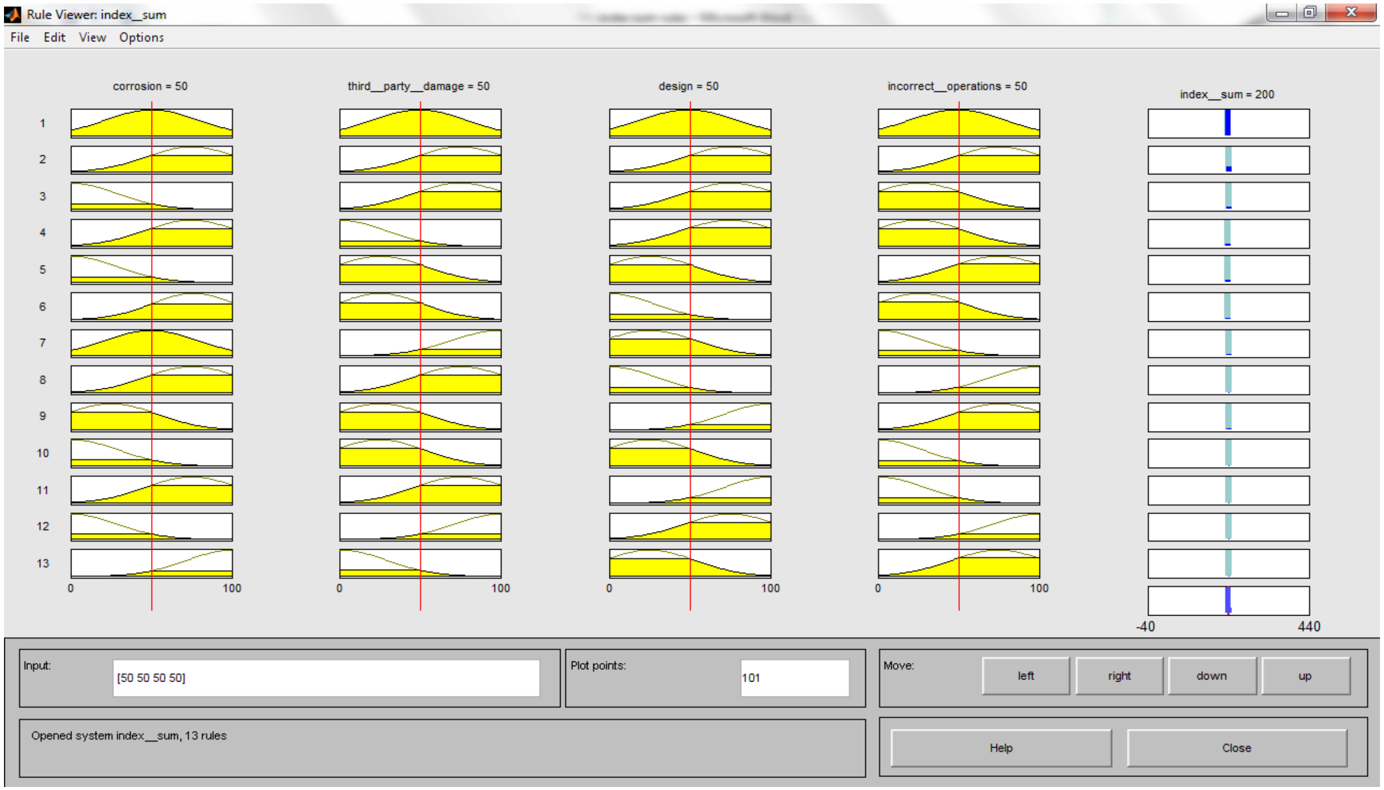


Fig.5. IS fuzzy inference structure

Fig.6. LIF fuzzy inference structure



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

Fig.7. Rules viewer and rules generated of IS model

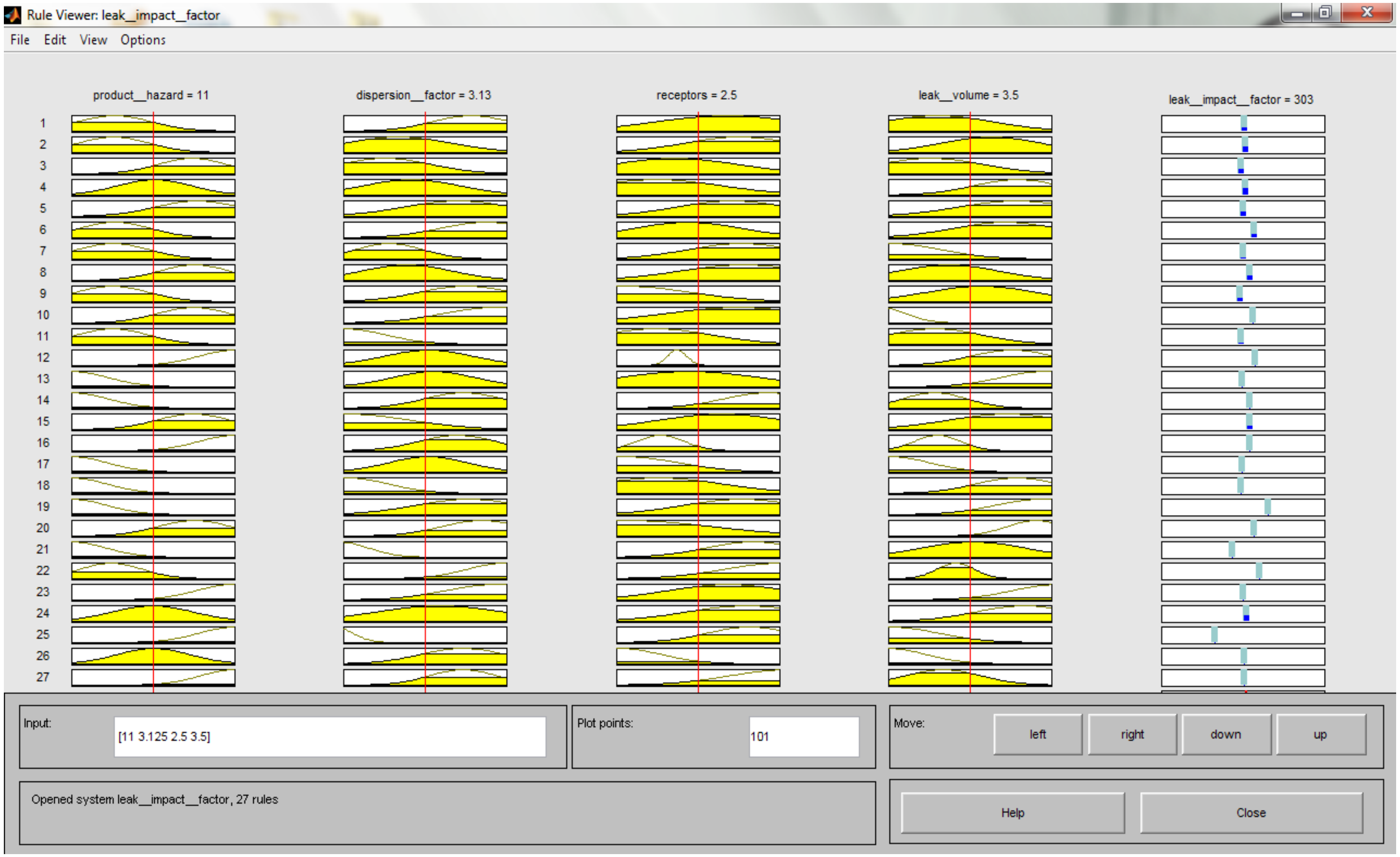


Fig.8. Rules viewer of LIF model

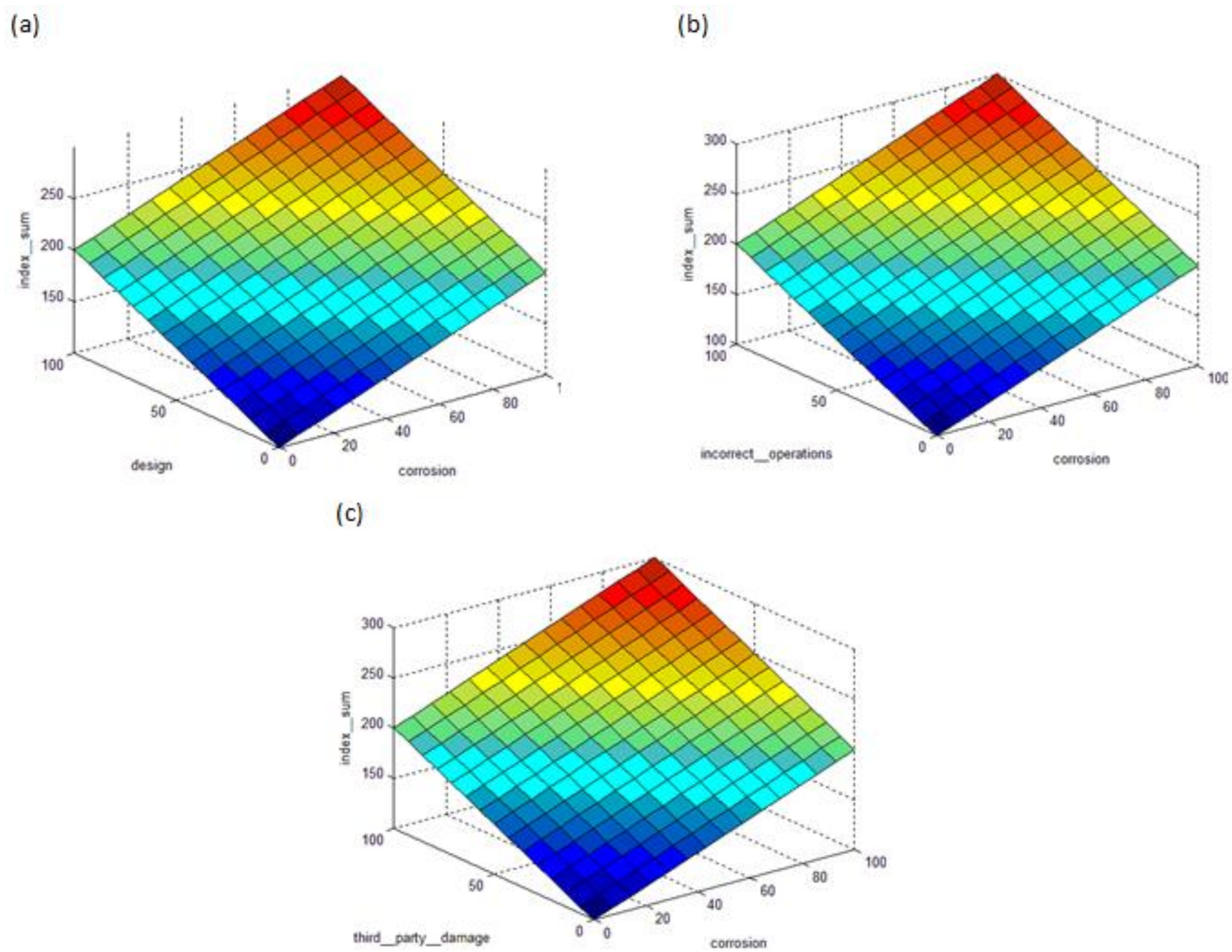


Fig.9. Control surface of IS on (a) design & corrosion; (b) incorrect operations & corrosion; (c) third party damage & corrosion

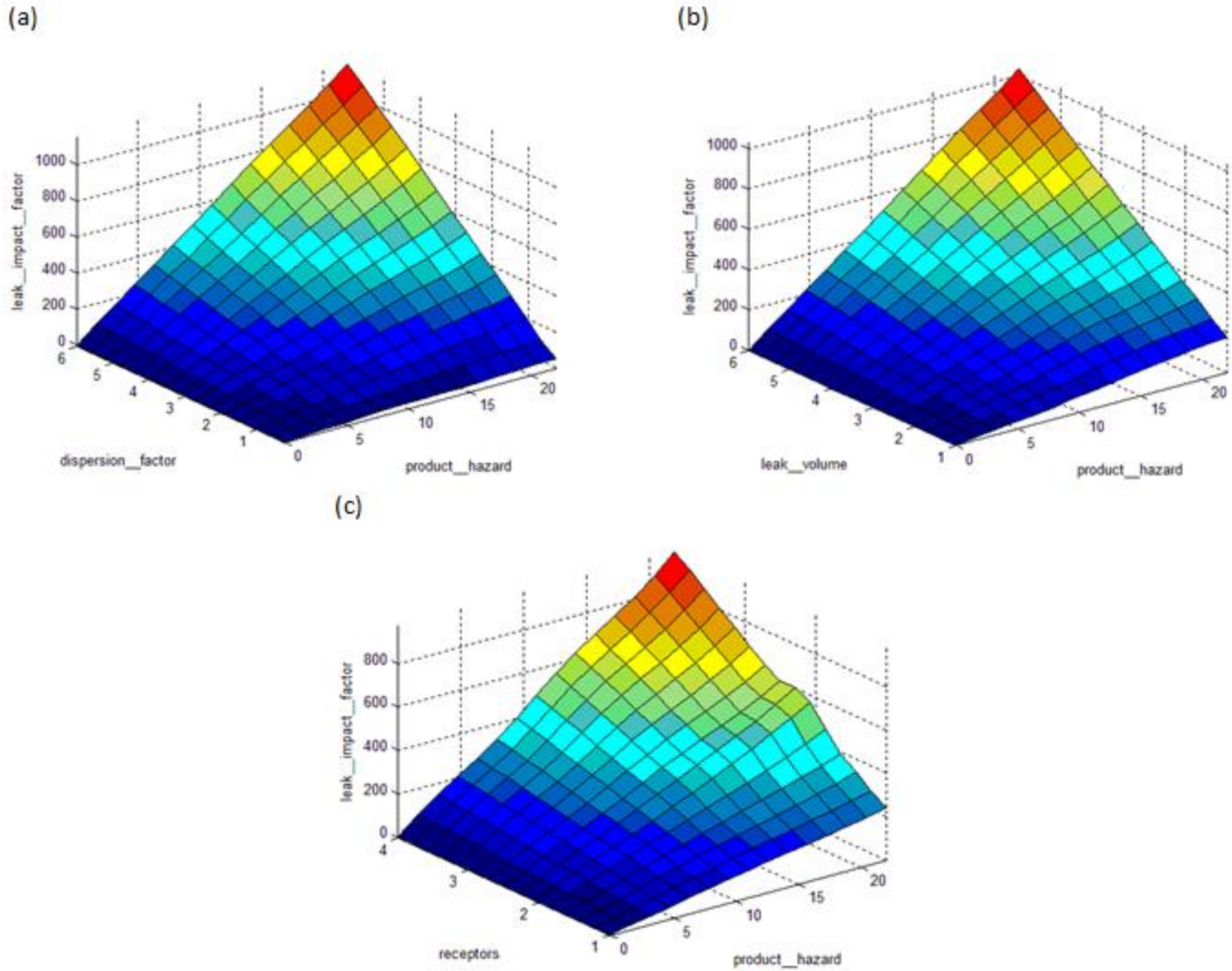


Fig.10. Control surface of LIF on (a) dispersion factor & product hazard; (b) leak volume & product hazard; (c) receptors & product hazard

8. Case study

- Crude oil is discharged from the VLCC's to the SUMED 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 where a pressure relief station preserves the pipeline from any over pressure.

- 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).

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 crops, high population density.
4	115 km – 165 km	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 crops, no ROW.
7	295 km – 320 km	25 km	Ends at Sidi Kerir, passing through lake, high population density.

Table.4. Pipeline sections & characteristics

- The risk assessment is performed by using the traditional method and by the proposed model on each pipeline section separately.

- 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:
- $IS_{(section1)} = 84 + 83 + 1 + 82 = 250$
- $DF_{(section1)} = 2/2 = 1$
- $LIF_{(section1)} = 9 \times 2 \times 1 \times 2 = 36$
- $RRS_{(section1)} = \frac{IS(section1)}{LIF(section1)} = 250/36 = 6.94$

9. Results and discussions

Section number	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

Table.5. Output RRS results of the proposed model

- 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.

- A comparison between traditional method and proposed model shows very high correlation and proves that the proposed model using fuzzy clustering is a powerful tool for pipe line risk assessment.

Section number	Traditional method						Proposed model					
	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	285.5	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

Table.6. Output RRS results of traditional method and proposed model

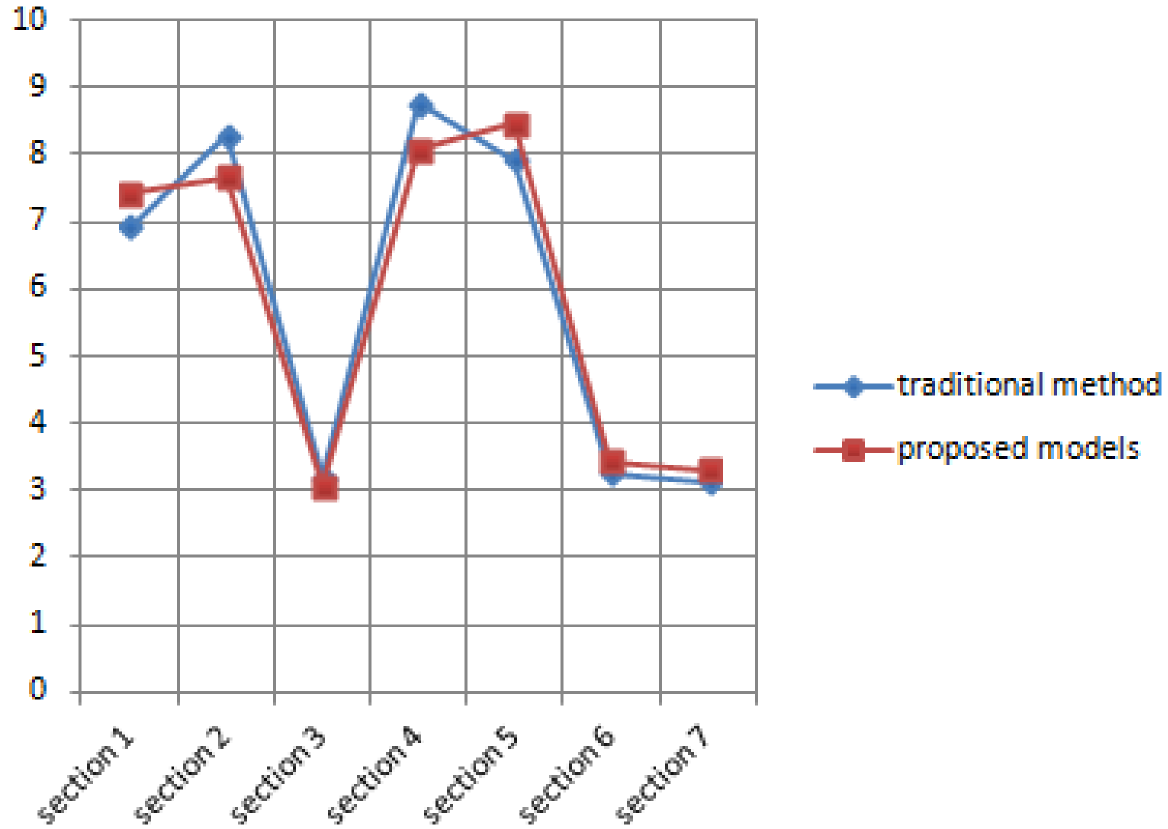


Fig.12. RRS results of traditional method versus proposed model

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

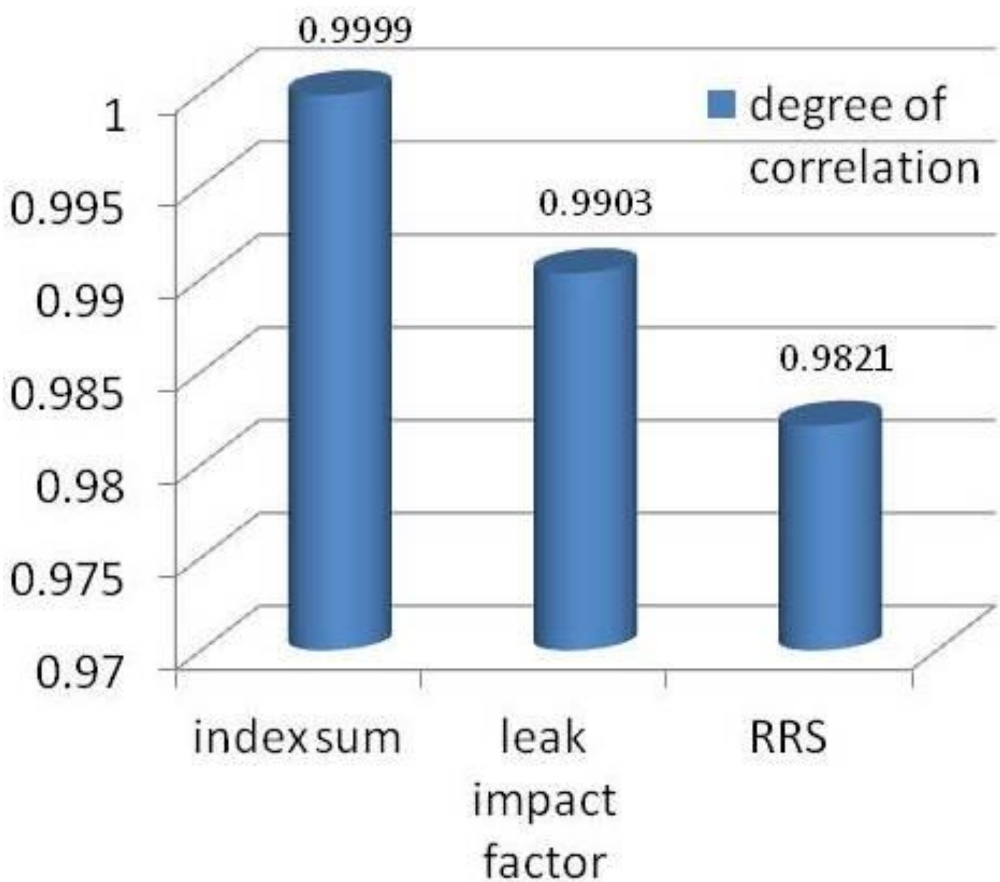


Fig.13. Correlation coefficient degree between qualitative and subtractive clustering model

10. 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.

- 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 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.