SOCAR Proceedings

Oil and gas structures and equipment

journal home page: http://proceedings.socar.az

INNOVATIVE METHODOLOGY FOR PREDICTING PIPE THICKNESS LOSS BASED ON ELECTROMAGNETIC FLAW DETECTION RESULTS

O. A. Dyshin, L. R. Bakirova, G. E. Feyziyeva*

Azerbaijan State Oil and Industry University, Baku, Azerbaijan

A B S T R A C T

The paper proposes an innovative approach to forecasting based on the results of electromagnetic flaw detection to determine maximum pipe thickness losses over 10-m intervals of the immersion depth of the lower part of a pipe into a well. The main model uses rank transformation applied to a fuzzy regression model with fuzzy input variables and fuzzy output. The input variables include the main parameters of the reservoir in the well (temperature, density, and dynamic viscosity of the hydrocarbon mixture), while the output variable is the maximum loss of pipe thickness in the above–mentioned immersion intervals of the lower part of the pipe. The error of the output forecasts is determined using a numerical procedure to estimate the difference between fuzzy numbers. The forecasting is performed using a sliding method that combines rank fuzzy regression with clear nonlinear regression until the prediction error reaches the specified threshold value. The use of fuzzy regression with fuzzy input and output makes it possible to assess the impact of reservoir parameters on the condition of pipes and the potential for emissions and critical situations.

Keywords: Fuzzy LR-type numbers; rank transformation; fuzzy regression; electromagnetic flaw detection; fuzzy outliers; sliding prediction; membership function; pipe thickness loss.

Introduction

Various geophysical investigation methods, including electromagnetic flaw detection (EMFD), thermometry, and noise metrics, are currently used to assess the technical condition of casing strings, drill pipes, and tubing [1, 2]. In the oil fields of Azerbaijan, the most commonly used method for evaluating the technical condition of wells with varying degrees of environmental aggressiveness is EMFD [3, 4].

In our study, we focused on wells in the Bulla-More field (Azerbaijan) at depths ranging from 1998.9 to 2381.5 m with different levels of aggressiveness. We employed EMFD type 43 to assess the condition of these wells, enabling us to determine pipe wall thickness, crack locations, and corrosion levels. Defects and damages were identified in technical strings with diameters D=127.0 mm, D=339.7 mm, and D=473.7 mm (where D represent the external and internal diameters of the string, respectively).

The purpose of this research is to develop a method for predicting the outer pipe thickness loss of the technical string for each 10-m depth interval of immersion into the well. This prediction is based on previous flaw detection measurements and takes into account possible changes in temperature, density, and viscosity of the fluid flow in the well at every 100-m depth interval.

**E-mail:* gulnarafeyziyeva5@gmail.com

http://dx.doi.org/10.5510/OGP20240100951

Research methodology

In order to solve the issues stated in the formulated research objective, the rank transformation in regression models with the advantages in the applications, which was first used by the authors [5] as an alternative procedure for analyzing experimental data, are significantly used. Fuzzy regression models, as well as ordinary regression models with deterministic independent variables, are sensitive to deviations and with a large number of initial data the estimates obtained with their help, have a wide scope [6, 7]. The method of rank transformation (RT-method) is one of the nonparametric methods of data analysis, independent of the form of data distribution and insensitive to deviations. In [8, 9] it is shown that the RT-method is stable and represents a high-power procedure for testing statistical hypotheses. The traditional fuzzy regression model with crisp coefficients was generalized in to account for cases in which regression coefficients for independent variables and the variables themselves are fuzzy numbers. This can be expressed in the following form:

$$Y_i\left(X_i\right) = A_0 + A_1X_{i1} + \ldots + A_pX_{ip} \quad (1)$$

where X_{ij}, A_j, and $Y_i(\boldsymbol{X}_i)$ are LR fuzzy numbers. The α-level set of the fuzzy regression model in Equation (1) can be expressed in the following form:

$$Y_i\left(\boldsymbol{X}_i\right) = \sum_{k=0}^{p}\left[m_l\left(\alpha\right), m_r\left(\alpha\right)\right] \quad (2)$$

where

$$m_l(\alpha)=\min\{l_k(\alpha)l_{X_{ik}}(\alpha),\ r_k(\alpha)l_{X_{ik}}(\alpha),\ l_k(\alpha)r_{X_{ik}}(\alpha),\ r_k(\alpha)r_{X_{ik}}(\alpha)\},$$

$$m_r(\alpha)=\max\{l_k(\alpha)l_{X_{ik}}(\alpha),\ r_k(\alpha)l_{X_{ik}}(\alpha),\ l_k(\alpha)r_{X_{ik}}(\alpha),\ r_k(\alpha)r_{X_{ik}}(\alpha)\},$$

$l_k(\alpha)$ and $r_k(\alpha)$ are left and right spreads of LR fuzzy numbers $A_{k,\alpha}=(a_k,l_k(\alpha),r_k(\alpha))_{LR}$ with mode and membership function as defined by Equation (5); $l_{X_k}(\alpha)$ and $r_{X_k}(\alpha)$ are left and right spreads of LR fuzzy numbers $X_{ik,\alpha}=(x_k,l_{X_{ik}}(\alpha),r_{X_{ik}}(\alpha))_{LR}$ with mode x_{ik} and membership function also defined by Equation (5).

If independent variables X_{ik} are positive fuzzy integers and $l_k(\alpha)\geq 0$ for each , then the α-level set for the fuzzy regression model can be written as follows [10]:

$$Y_i(\boldsymbol{X}_i)(\alpha)=\sum_{k=0}^{p}\left[l_k(\alpha)l_{X_{ik}}(\alpha),\ r_k(\alpha)r_{X_{ik}}(\alpha)\right] \quad (3)$$

A LR fuzzy number $A=(a,l,r)_{LR}$ is defined as a fuzzy number A with a membership function $\mu_A(x)$, defined in the following form:

$$\mu_A(x)=\begin{cases}L_A\left((a-x)/l\right), if\ 0\leq a-x\leq l,\\ R_A\left((x-a)/r\right), if\ 0\leq x-a\leq r,\end{cases} \quad (4)$$

where represents fuzzy number mode A, while a l and r represent the left and right spreads (or fuzzy parameters) of the fuzzy number A (fig. 1). In this case, $L_A((a-x)/l)$ and $R_A((a-x)/r)$ are referred to as the left and right branches of the membership function $\mu_A(x)$. Function $L_A(.)$ from argument $y((a-x)/l)$ decreases at yє [0,1] because $y=0\rightarrow x=a\rightarrow \mu(a)=1$ u $y=1\rightarrow x=a\rightarrow \mu(x)=0$; however, function $R_A(.)$ from argument $y=(x-o)/r$ decreases at yє [0,1] because $y=0\rightarrow x=a\rightarrow \mu(a)=1$ u $y=1\rightarrow x=a+r\rightarrow \mu(x)=0$. Consequently,

$$L_A(0)=R_A(0)=1 \text{ u } L_A(1)=R_A(1)=0 \quad (5)$$

Here, $L_A^{-1}(\alpha)$ and $R_A^{-1}(\alpha)$ are the points on the x-axis. Located from the mode a of fuzzy number $A=(a,l,r)_{LR}$, at distances of $l_A(\alpha)$ and $r_A(\alpha)$, are abscissas of the points of intersection of the line $\mu_A(x)=\alpha$ with the left and right spreads of the membership function $\mu_A(x)$ of the fuzzy number A, respectively. When $x=L_A^{-1}(\alpha)$ and $R_A^{-1}(\alpha)$, the inverse functions $L_A(.)$ and $R_A(.)$ i.e., $L_A^{-1}(\alpha)$ and $R_A^{-1}(\alpha)$, are solutions to the equations $L_A(\frac{a-x}{l})=\alpha$ and $R_A(\frac{x-a}{l})=\alpha$, respectively.

For the LR fuzzy number $A=(a,l,r)_{LR}$, we introduce the notion of an α-level set defined by the following relation [10]:

$$A(\alpha)=\begin{cases}\{x:\mu_A(x)>\alpha\}, if\ \alpha=0,\\ \{x:\mu_A(x)\geq\alpha\}, if\ 0<\alpha\leq 1,\end{cases} \quad (6)$$

where, for any set A, the symbol $\bar{A}$ denotes the closure of the set A. The α-level set $A(\alpha)$ for the LR fuzzy number $A=(a,l,r)_{LR}$ is a closed interval of the following form:

$$A(\alpha)=[a-l_A(\alpha),\ a+r_A(\alpha)] \quad (7)$$

The α-level set of the observed fuzzy number $Y_i=(\hat{y}_i,l_{\hat{y}_i}(0),\ r_{\hat{y}_i}(0))_{LR}$ in Equation (1) can be written as a product of intervals

$$\sum_{k=0}^{p}\left[l_{A_k}(\alpha),r_{A_k}(\alpha)\right]\cdot\left[l_{X_{ik}}(\alpha),r_{X_{ik}}(\alpha)\right], \quad (8)$$

where $l_{A_k}(\alpha)=a_k-L_{A_k}^{-1}(\alpha),\ r_{A_k}(\alpha)=R_{A_k}^{-1}(\alpha)-a_k$

$l_{X_{ik}}(\alpha)=\hat{x}_{ik}-L_{X_{ik}}^{-1}(\alpha),\ r_{X_{ik}}(\alpha)=R_{X_{ik}}^{-1}(\alpha)-\hat{x}_{ik}x$.

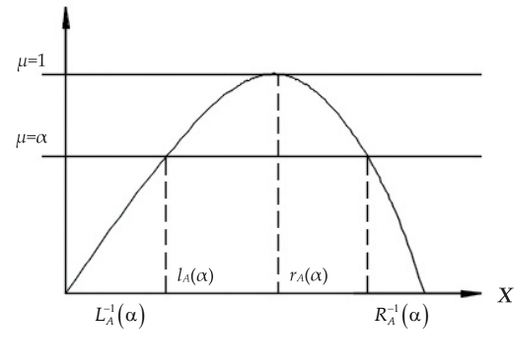

Fig. LR fuzzy number $A=(a,l,r)$ with mode a and spreads l and r

The ranks of crisp numbers are defined as follows.

Let $X_1,\dots,A_n$ be a sequence of independent crisp numbers (e.g., observations of the same variable X) and $X_{(1)}\leq X_{(2)}\leq\dots\leq X_{(n)}$ be its corresponding ordinal series of numbers, i.e., a series of the same numbers but ordered in the non-decreasing order. The first number in the ordered series is assigned a rank $R=1$, second one $R=2$, etc.; in the case of two identical numbers equal to r, for example, in the unordered series, the first one is assigned a rank $R=[r-1+r]/2$, and the second one rank $R=[r+(r+1)]/2$. For an arbitrary number X_{n+1} that may not be included in the original series of numbers $\{X_i\}$, the rank is calculated according to the following rules [7].

1. If $X_{n+1}<X_{(1)}$, then $R(X_{n+1})=R(X_{(1)})$
2. If $X_{n+1}>X_{(n)}$, then $R(X_{n+1})=R(X_{(1n)})$
3. If $X_{n+1}=X_{(i)}$ for some i, then $R(X_{n+1})=R(X_{(i)})$ (9)
4. If $X_{(i)}<X_{n+1}<X_{(i+1)}$, $i=1,2,\dots,n-1$ then

$R(X_{n+1})=R(X_{(i)})+[R(X_{i+1})-R(X_i)]\cdot[X_{n+1}-X_{(i)}]-[X_{(i+1)}-X_{(i)}]$

In the case of crisp regression, i.e., when the vector $X_{(ik)}$ are observations of independent variables X_k ($k=1,\dots,p, i=1,\dots,n$), predicted ranks $\hat{R}(Y_i)$ are expressed through ranks $R(X_{ik})$. For simple regression Y from X ($p=1$), the rank transformation (RT) method yields the following equation:

$$R(Y_i)=(n+1)/2+\beta_1[R(X_i)-(n+1)/2] \quad (10)$$

where β_1 is the Spearman rank correlation coefficient.

In this case, the algorithm for calculating values $Y_i(\boldsymbol{X}_i)(\alpha)$ consists of the following steps [8].

1. We calculate rank $R(x_{ik})$ of observations x_{ik} from the sequence of observations $\{x_{1k},\dots,x_{nk}\}$ and rank $R(l_{Y_i}(\alpha^*))$ of spreads $l_{Y_i}(\alpha^*)$ by the rules presented in Equation (9) when $\alpha^*\in[0,1]$.

2. Based on the sample

$\{(R(x_{i1}),\dots,(R(x_{ip}),R(l_{Y_i}(\alpha^*)):i=1,\dots,n\}$, we can construct a regression model:

$$R\left(l_{Y_i}(\alpha^*)\right)=\beta_0(\alpha^*)+\sum_{k\subset 1}^{p}\beta_k(\alpha^*)R(x_{ik})+\sum_{i=1}^{p}\sum_{k=1}^{p}\beta_{kj}(\alpha^*)R(x_{ik})R(x_{ij}) \quad (11)$$

3. Using the LMS or LAD method, we calculate the ranks $\hat{R}(l_{Y_i}(\alpha^*))$ predicted by the model (Equation (12)) and a set of ranks $\{R(x_{ik})\}$ of observed values of the independent variables.

4. We calculate the values $\bar{l}_{\hat{Y}_i}(\alpha^*)$ according to the following rules:

$$\overline{l}_{\hat{Y}_i}(\alpha^*)=\begin{cases} l_{Y_{(1)}}(\alpha^*), \text{ if } \hat{R}\left(l_{Y_i}(\alpha^*)\right)<R\left(l_{Y_{(1)}}(\alpha^*)\right), \\ l_{Y_{(n)}}(\alpha^*), \text{ if } \hat{R}\left(l_{Y_i}(\alpha^*)\right)>R\left(l_{Y_{(n)}}(\alpha^*)\right), \\ l_{Y_{(j)}}(\alpha^*), \text{ if } \hat{R}\left(l_{Y_j}(\alpha^*)\right)=R\left(l_{Y_{(j)}}(\alpha^*)\right). \end{cases} \quad (12)$$

If $R(l_{Y(j)}(\alpha^*))<\hat{R}\,(l_{Y(j)}(\alpha^*))<R(l_{Y(j+1)}(\alpha^*))$, then

$$\overline{l}_{\hat{Y}_i}(\alpha^*)=l_{Y_{(j)}}(\alpha^*)+\left(l_{Y_{(j+1)}}(\alpha^*)-l_{Y_{(j)}}(\alpha^*)\right)\times \times\frac{R\left(l_{Y_{(i)}}(\alpha^*)\right)-R\left(l_{Y_{(j)}}(\alpha^*)\right)}{R\left(l_{Y_{(j+1)}}(\alpha^*)\right)-R\left(l_{Y_{(j)}}(\alpha^*)\right)} \quad (13)$$

where $l_{Y(j)}(\alpha^*)$ is the j-e ascending value in the sequence $\{l_{Y_1}(\alpha^*), \dots, l_{Y_n}(\alpha^*)\}$. As the left-hand side value of the α-level set of the fuzzy number should be no more than its mode, the estimate of the left spread of the fuzzy number $Y_i(\boldsymbol{X}_i)(\alpha^*)$ can be written as

$$\hat{l}_{\hat{Y}_i}(\alpha^*)=\min\left\{\overline{l}_{\hat{Y}_i}(\alpha^*),\hat{y}_i\right\} \quad (14)$$

where $\hat{y}_i$ is the a priori estimation of the fuzzy number mode $\hat{y}_i\,(\alpha)$, which can be taken as an estimate $\hat{a}_{Y_i}^{(v_0)}$ of the modes of the fuzzy number Y_i, obtained by fuzzification according to Equation (26).

Similarly, for the sample $\{(R(x_{i1}),\dots,(R(x_{ip}), R(r_{Y_i}(\alpha^*)):i=1, \dots,n\}$, by substituting l to r in Equations (12)–(13), we can obtain an estimate of the right spread of the fuzzy number $Y_i(\alpha^*)$ as follows:

$$\hat{r}_{\hat{Y}_i}(\alpha^*)=\max\left\{\overline{r}_{\hat{Y}_i}(\alpha^*),\hat{y}_i\right\} \quad (15)$$

Let us now construct a parametric fuzzy regression model to estimate the spreads of the α-level fuzzy number set $Y_i(\boldsymbol{X}_i)(\alpha)$ when $\alpha\neq\alpha^*$. Using the estimates $\hat{r}_{\hat{Y}_i}(\alpha^*)$ and $\overline{l}_{\hat{Y}_i}(\alpha^*)$ obtained by steps 1–4, the right and left spreads at each value of the parameter α are determined by the following formula:

$$\hat{r}_{\hat{Y}_i}(\alpha)=\begin{cases} \max\left\{\max\limits_{\{\alpha\le s<\alpha^*\}}\left\{\overline{r}_{Y_i}(s)\right\},\hat{y}_i\right\}, \text{ if } a<\alpha^*, \\ \max\left\{\min\limits_{\{\alpha^*<s\le\alpha\}}\left\{\overline{r}_{Y_i}(s)\right\},\hat{y}_i\right\}, \text{ if } \alpha^*<a \end{cases}$$

and (16)

$$\hat{l}_{\hat{Y}_i}(\alpha)=\begin{cases} \min\left\{\max\limits_{\{\alpha^*\le s<\alpha\}}\left\{\overline{l}_{Y_i}(s)\right\},\hat{y}_i\right\}, \text{ if } \alpha^*<a, \\ \min\left\{\min\limits_{\{\alpha<s\le\alpha^*\}}\left\{\overline{l}_{Y_i}(s)\right\},\hat{y}_i\right\}, \text{ if } a<\alpha^* \end{cases}$$

The membership function of the estimated fuzzy output Y_i м can be determined by the parametric estimation method. For each fixed number $i(i=i_0-8,\dots,i_0+1)$, we use the given data samples $\{(\hat{l}_{\hat{Y}_i}(\alpha_k),\alpha_k):k=0,1,\dots,k_0\}$ and $\{(\hat{r}_{\hat{Y}_i}(\alpha_k),\alpha_k):k=0,1,\dots,k_0\}$ with the control increasing sequence $\{\alpha_k\}$ of the parameter values α (e.g., $\alpha_k=k/10$, $k=0,1,\dots,10$) and values $\hat{l}_{\hat{Y}_i}(\alpha_v)$ and $\hat{r}_{\hat{Y}_i}(\alpha_v)$, as defined by Equation (17) to construct regression equations:

$$R_{\hat{Y}_i}(y)=\beta_0^{(r)}+\beta_1^{(r)}\cdot y+\beta_2^{(r)}\cdot y^2 \quad (17)$$

and

$$L_{\hat{Y}_i}(y)=\beta_0^{(l)}+\beta_1^{(l)}\cdot y+\beta_2^{(l)}\cdot y^2 \quad (18)$$

The values of $\hat{R}_{\hat{Y}_i}(y)$ and $\hat{L}_{\hat{Y}_i}(y)$ predicted by regression (Equations (14) and (15)) are the estimates of the right branch of $R_{\hat{Y}_i}(y)$ and the left branch of $L_{\hat{Y}_i}(y)$ of the membership function $\mu_{\hat{Y}_i}(y)$. Furthermore, by denoting the obvious mode estimates of the fuzzy number $\hat{Y}_i$ as $\hat{y}_{\hat{Y}_i}^{(l)}=\hat{l}_{\hat{Y}_i}(1)$ and $\hat{y}_{\hat{Y}_i}^{(r)}=\hat{r}_{\hat{Y}_i}(1)\,\hat{Y}_i$, the estimate of the fuzzy number $\hat{y}_i$ can be determined as follows:

$$\hat{y}_i=\frac{(\hat{y}_{\hat{Y}_i}^{(l)}+\hat{y}_{\hat{Y}_i}^{(r)})}{2} \quad (19)$$

The spreads $l_{\hat{Y}_i}$ and $r_{\hat{Y}_i}$ of the fuzzy number $\hat{Y}_i$ are estimated as

$$\hat{l}_{\hat{Y}_i}=\hat{y}_i-L_{\hat{Y}_i}^{-1}(0),\ \hat{r}_{\hat{Y}_i}=R_{\hat{Y}_i}^{-1}(0)-\hat{y}_i \quad (20)$$

where $L_{\hat{Y}_i}^{-1}(0)$ and $R_{\hat{Y}_i}^{-1}(0)$ are the solutions of equations $L_{\hat{Y}_i}(y)=0$ and $R_{\hat{Y}_i}(y)=0$, respectively. Consequently, the fuzzy number $\hat{Y}_i$ is represented in the LR form as

$$\hat{Y}_i=\left(\hat{y}_i,\hat{y}_i-L_{\hat{Y}_i}^{-1}(0),R_{\hat{Y}_i}^{-1}(0)-\hat{y}_i\right)_{LR} \quad (21)$$

Let us now consider the fuzzy regression presented in Equation (1) in the fuzzy input/output case. Using sample data $\{(l_{X_{i1}}(\alpha),\dots,l_{X_{ip}}(\alpha),l_{Y_i}(\alpha)):i=1,\dots,n\}$, we construct a rank regression as

$$R\left(l_{Y_i}(\alpha^*)\right)=\beta_0^{l}(\alpha^*)+\sum_{k=1}^{p}\beta_k^{(l)}(\alpha^*)R\left(l_{X_{ik}}(\alpha^*)\right)+ +\sum_{j=1}^{p}\sum_{k=1}^{p}\beta_{kj}^{(l)}(\alpha^*)R\left(l_{X_{ik}}(\alpha^*)\right)\cdot R\left(l_{X_{ij}}(\alpha^*)\right) \quad (22)$$

In addition, based on sample data $\{(r_{X_{i1}}(\alpha),\dots,r_{X_{ip}}(\alpha), r_{Y_i}(\alpha)):i=1,\dots,n\}$, we construct another rank regression as

$$R\left(r_{Y_i}(\alpha^*)\right)=\beta_0^{r}(\alpha^*)+\sum_{k=1}^{p}\beta_k^{(r)}(\alpha^*)R\left(r_{X_{ik}}(\alpha^*)\right)+ +\sum_{j=1}^{p}\sum_{k=1}^{p}\beta_{kj}^{(r)}(\alpha^*)R\left(r_{X_{ik}}(\alpha^*)\right)\cdot R\left(r_{X_{ij}}(\alpha^*)\right) \quad (23)$$

The estimates of the output spreads $Y_i(\boldsymbol{X}_i)(\alpha)$ can now be obtained using Equation (16), and the mode estimate can be obtained using Equation (21).

Application of rank-based fuzzy regression to predicting flaw detection measurements

Example. Let us analyze the flaw detection measurement data presented in table 1, sourced from gas reporting documentation of the geophysical department of the oil company SOCAR in Azerbaijan.

Let us denote the output variable as Y. This represents the maximum loss of pipe thickness (in %) (column 7 of table 1) at each approximately 10-m depth of immersion of the lower part of the outer pipe into the borehole (column 2). The he depth of immersion of the lower part of the pipe into the borehole is represented by i values in column 2 of table 1. Each i corresponds to a values Y_i of the variable Y. Let i_0 be the last known observation from column 2 with the corresponding observation Y_i from column 7. The task is to predict the variable Y at depths $i>i_0$ with a step of 10 m.

The main input variables affecting the output variable Y are the following reservoir parameters in the borehole: temperature T, °C; fluid density ρ, kg/m^3 dynamic viscosity ν, poise. These parameters are denoted as X_1, X_2 and X_3. Data for these parameters is presented in table 2 annually from 2019 to 2022, covering depths j ranging from 0 to 2500 m in 100 m increments. It is assumed that these parameters remain relatively constant within each 100 m depth interval.

Consequently, if the depth from column 2 of table 1 is related to the depth j from column 2 of table 2 by the inequality $j<i\le j+1$ ($j=2000,\dots,2500$), then $X_{ik}=X_{jk}$, ($k=1, 2, 3$), where $k=1, 2, 3$ is variable of the number X.

Each crisp value X_k can be made fuzzy (i.e., fuzzified)

Table 1

Dependence of maximum pipe thickness loss on the depth of immersion in the borehole

Upper part of the pipe (m)	Bottom part of the pipe (m)	Pipe length (m)	Nominal thickness of the pipe (mm)	Actual minimum thickness of the pipe (mm)	Depth of maximum pipe thickness losses (m)	Maximum loss of pipe thickness (%)	Classification of losses
1	2	3	4	5	6	7	8
2283.10	2292.40	8.29	13.07	12.40	2285.10	5.1	B
2292.70	2302.00	8.29	13.07	12.52	2300.00	4.2	A
2301.30	2311.60	8.34	13.07	12.13	2308.30	7.1	B
2311.90	2321.20	8.29	13.07	12.25	2321.20	6.1	B
2321.40	2330.80	8.34	13.07	12.17	2323.50	6.8	B
2331.10	2340.60	8.50	13.07	12.42	2335.70	5.0	A
2341.90	2350.00	8.11	13.07	12.56	2347.40	3.7	A
2351.30	2359.20	7.98	13.07	8.55	2359.20	34.6	E
2361.20	2369.10	7.86	13.07	8.35	2366.80	36.1	E
2370.40	2381.00	10.61	13.07	8.55	2380.70	34.6	E

using fuzzy regression (Equation (1)), with p=1, Y=X_k, and X=j. The membership function of the LR fuzzy number is calculated using the above method of fuzzy regression for each X_{ik} at any value $i \le i_0$, based on the values in column 2 of table 2 and the estimates of the mode and spreads of the fuzzy number X_k.

Assuming that the variables X_1, X_2, X_3, and Y (denoted as (c_k)) are fuzzy normal variables following the definition presented by Nahmias [9], having distribution possibilities with center c_k and spread w_k:

$$\pi_{C_k}(q) = \exp\left(-\frac{(q-c_k)^2}{w_k}\right) \tag{24}$$

In order to estimate the parameters c_k, w_k of the distribution in Equation (24), we use the results of Cai [10], which were obtained for normal fuzzy variables as defined above.

For a normal fuzzy variable on the image space (Γ, У, σ), we have

$$\mu_X(x) = \sigma(X = x) = \exp\left[-\left(\frac{x-a}{b}\right)^2\right] \tag{25}$$

Let X_j be the j^{th} observation of variable X at time t_j and $X_1, \ldots, X_N$ be independent variables obeying the normal law $N(a, b)$ with $x_1, \ldots, x_N$ being their realization. Given a possibility α, we can define ε_α such that

$$\sigma\left(\left|\frac{X_j - a}{b}\right| > \varepsilon_\alpha, j = 1, \ldots, N\right) = \min_{1 \le j \le N} \sigma\left(\left|\frac{X_j - a}{b}\right| > \varepsilon_\alpha\right) = \alpha$$

Then, with risk α, the following parameter estimates are valid for a and b [11]:

$$\begin{aligned} \hat{a} &= \frac{1}{2}\left(\max_{1 \le j \le N} x_j + \min_{1 \le j \le N} x_j\right) \\ \hat{b} &= \left(\max_{1 \le j \le N} x_j - \min_{1 \le j \le N} x_j\right) / 2\varepsilon_\alpha \end{aligned} \tag{26}$$

The calculations are based on $\alpha = \overset{0}{\alpha} = e^{-4}$; hence, $\varepsilon_{\overset{0}{\alpha}} = 2$.

We denote the intervals $\hat{Y}_i(0)$ and $Y_i(0)$ as

$$\hat{Y}_i(0) = \left[\hat{y}_i - \hat{l}_{\hat{Y}_i}(0),\ \hat{r}_{\hat{Y}_i}(0) - \hat{y}_i\right]$$

and

$$Y_i(0) = \left[\hat{a}_{Y_i}^{(v_0)} - L_{Y_i}^{-1}(0),\ R_{Y_i}^{-1}(0) - \hat{a}_{Y_i}^{(v_0)}\right]$$

Table 2

Dependence of reservoir parameters (density ρ, kg/m³ and viscosity ν, poise) from the depth of immersion of the pipe into the borehole from 2019 to 2022

Years							
2019		2020		2021		2022	
ρ	ν	ρ	ν	ρ	ν	ρ	ν
896	6.23	898	5.70	890	5.51	890	5.10
890	6.20	897	5.68	890	5.50	890	5.05
889	6.17	895	5.66	890	5.47	888	5.00
889	6.14	892	5.60	886	5.45	886	4.98
888	6.13	890	5.53	882	5.41	884	4.95
822	6.10	886	5.51	880	5.40	882	4.90
891	6.05	885	5.51	876	5.37	880	4.87
890	6.03	880	5.51	874	5.33	878	4.82
889	5.95	878	5.45	866	5.26	863	4.76
868	5.93	875	5.40	863	5.06	860	4.72
885	5.90	871	5.33	880	5.16	850	4.69
886	5.87	870	5.30	870	5.10	845	4.67
882	5.81	866	5.27	871	5.06	840	4.65
880	5.78	862	5.25	870	5.01	834	4.61
874	5.73	860	5.22	866	5.00	830	4.58
870	5.71	855	5.20	863	4.90	818	4.52
869	5.73	851	5.18	860	4.85	815	4.48
868	5.70	848	5.16	858	4.81	810	4.40
860	5.65	842	5.15	856	4.72	810	4.31
859	5.63	840	5.14	857	4.70	808	4.27
855	5.60	840	5.14	850	4.60	807	4.25
853	5.60	835	5.13	848	4.62	805	4.25

The measure of difference between fuzzy numbers Y_i and $\hat{Y}_i$ is defined by the following formula [8].

In order to compare the estimates of the output variable Y_i predicted by the fuzzy regression presented in Equation (1) with its actual values, we use a method to detect the difference between the fuzzy numbers Y_i and $\hat{Y}_i$ [8]:

$$d\left(Y_i,\hat{Y}_i\right)=\frac{\int_{-\infty}^{\infty}\left|\mu_{Y_i}(x)-\mu_{\hat{Y}_i}(x)\right|dx}{\int_{-\infty}^{\infty}\mu_{Y_i}(x)dx}+h_d\left(Y_i(0),\ \hat{Y}_i(0)\right) \qquad (27)$$

where $h_d(Y_i(0),\ \hat{Y}_i(0))$ is the relative difference between the intervals $Y_i(0)$ and $\hat{Y}_i(0)$, which is estimated as the ratio of the distance $\rho_d(Y_i(0),\ \hat{Y}_i(0))$ to the length of the interval $Y_i(0)$, with $\rho_d(Y_i(0),\ \hat{Y}_i(0))$ being the distance between the intervals $Y_i(0)$ and $\hat{Y}_i(0)$.

Let us consider the problem of predicting the values of the output variable using the example of flaw detection measurements given in tables 1 and 2.

Suppose we only have observations from table 1 for a certain depth i_0 of pipe immersion in the borehole, satisfying the condition $j_0<i_0\le j_0+1$ ($j_0=2000,\ldots,2400$), where j is the depth of the borehole from column 2 of table 2, and we need to calculate the predictions at each new 10-m mark of the depth i of pipe immersion in the borehole. To this end, we propose an algorithm based on the rank transformation of fuzzy regression and the method of estimating the best polynomial regression $y=f(x)$.

Algorithm for predicting the maximum loss of pipe thickness by the depth of immersion into the borehole

1. For each parameter X of the reservoir, we introduce the notations $y=X$, $x=h$ (h= borehole depth). Using the data from table 3, we construct a sequence of two-dimensional points (x_i, y_i), ($i=0,1,\ldots,m-1$; $m=6$), $y_i=y(x_i)$; $x_i=2000+i\cdot\Delta x$, ($\Delta x=100$; $i=0,1,\ldots,5$). By sampling $V=\{x_i, y_i\}$ ($j=0,1,\ldots,5$) using the least squares (LS) method, we construct a regression model as a polynomial of degree $l\le4$:

$$y=f(x)=P_l\left(x\right)=\sum_{v}^{l}a_v x^v$$

2. Within each interval $[x_{j-1}, x_j]$ ($j=1,\ldots,n$) ($n=5$), we select an interpolation node $x_j^*=x_j-\frac{10}{3}$. Based on points $\{x_j^*\}$ ($j=1$, $\ldots n$), we construct an interpolation Lagrange polynomial [12]:

$$g_n(x)=L_n(x)=\sum_{i=1}^{n}f(x_i^*)\cdot\prod_{j\ne i}\frac{x-x_j^*}{x_i^*-x_j^*} \qquad (28)$$

Since $f^{(n)}(x)\equiv0$ (as a result of $l\le4$), we have $f(x)\equiv g_n(x)$ for any $x\in[x_0, x_5]$, satisfying the condition $x\ne x_j^*$ ($j=1,\ldots,n$), particularly for all $x_{\tilde{j}}=2000+\tilde{j}\cdot\Delta\tilde{x}$ ($\Delta\tilde{x}=10;\tilde{j}=0,1,\ldots,50$).

Thus, using the known values of variables X on the 100-m depth scale h in the interval [2000; 2500], their values on the 10-m depth scale h in the interval [2000; 2500] can be calculated using the interpolation Lagrange polynomial.

3. Using h_{i_0} we denote the depth of pipe immersion into the borehole in a 10-m scale, at which the flaw detection measurements were stopped. Subsequently, we use substitutions $\tilde{h}_i=h_i/2000$ and $\tilde{h}_i=h_i/2000$.

4. Assuming $i=1,2,\ldots$ and interval lengths $\Delta i=0.005$ m in the depth interval range [1; 1.25], let i_0 represent the number of the last available length interval Δi in the depth interval $[1;\tilde{h}_{i_0}]$. $\{y_i\}$, $i\in I=\{1,\ldots,i_0\}$ is a sequence of data in column 7 of table 1 on rows $i=1,\ldots,i_0$ (starting from the first row of column 7), characterizing the maximum losses (in %) at these values of i.

Let i_0 be the number of the last available interval, length $\Delta i=0.005$ m, and y_n be the maximum loss of pipe thickness defined by column 7 of table 1.

5. Based on the sample $V_n=\{x_i, y_i\}$ ($i=1,\ldots,i_0$), we construct the best polynomial regression model $y=F(x)$ of grades $l\le10$ using the LS method. We denote $f(x)=F'(x)$ as the function derivative of $F(x)$. We then apply the four-stage Runge–Kutta method of the fourth order with step $\Delta\tilde{h}=0.005$, assuming n_0 to be a crisp number (otherwise $\{V_n\}$ is considered, starting with c $n=2$):

$$y_{n+1}=y_n+\frac{1}{6}(k_1+2k_2+2k_3+k_4),$$
$$k_1=\Delta\tilde{h}\cdot f(x_n),\ k_2=\Delta\tilde{h}\cdot f\left(x_n+\frac{1}{2}\widetilde{K}_1\right) \qquad (29)$$
$$k_3=\Delta\tilde{h}\cdot f\left(x_n+\frac{1}{2}K_2\right),\ k_4=\Delta\tilde{h}\cdot f\left(x_n+K_3\right)$$

Using Equation (29), and knowing that $y_n=F(x_n)$ when $n=n_0$, the value of $y_{n_0+1}=F(x_{n_0+1})$. The forecast error ρ_1 can be estimated by the following equation [12] :

$$\rho_1=\frac{1}{2880}\left|f^{(4)}\left(x_{n_0}\right)\right|\cdot\Delta\tilde{h}^5+O\left(\Delta\tilde{h}^6\right) \qquad (30)$$

where $f^{(4)}(x_{n_0})$ is the fourth function derivative $f(x)$ on the point x_{n_0}.

By shifting the sample $\{V_n\}$ before by step h (assuming the obtained forecast value y_{n_0+1} for the actual value), we obtain the prediction value $y_{n_0+2}=F(x_{n_0+2})$ with a forecast error ρ_2, calculated by Equation (30) after substituting x_{n_0} in the right-hand side with x_{n_0+1}. In this way, it is possible to calculate forecasts up to a point $x=\tilde{h}_{\max}=1.25$ ($i_{\max}=50$).

6. We fuzzify the variable Y by its values y_i($j_0-1\le i\le i_0$), using the results of Nahmias [9] and Hong [10] on normal fuzzy variables.

$$\mu_Y(y)=\exp\left[-\left(\frac{y-a}{b}\right)^2\right]\sigma\left(\left|\frac{y_i-a}{b}\right|>\varepsilon_\alpha, i=1,\ldots,n\right)=\alpha \qquad (31)$$

where $\mu_Y(y)$ is the membership function of a fuzzy number Y; σ is the sign of a probability measure; $\alpha=\overset{0}{\alpha}=e^{-4}$, $\varepsilon_{\alpha}=2$.

With risk α, the parameter estimates are valid a and b [11]:

$$\hat{a}^{(0)}=\frac{1}{2}\left(\max_{i\in I_0}y_i+\min_{i\in I_0}y_i\right),\ \hat{b}^{(0)}=\left(\max_{i\in I_0}y_i-\min_{i\in I_0}y_i\right)/2\varepsilon_\alpha \qquad (32)$$

If $\hat{a}^{(0)}<y_{i_0}$, the sequence I_0 is replaced by a sequence shifted by 1, $I_0=\{2,3,\ldots,i_0,i_0+1\}$, and using Equation (32) for $i\in I_1$, new estimates are calculated as $\hat{a}^{(1)}$, $\hat{b}^{(1)}$ values for a and b, until at some step of v_0, the condition $\hat{a}^{(v_0)}\approx y_{i_0}$ and $\mu_Y(y_{i_0+1})\approx1$.

In this case, the function $\mu_Y(y)$ can be represented as a sequence of functions $\mu_Y^{(v)}(y)$ ($v=0,1,2,\ldots$) such that $\mu_Y^{(v_0)}(y_{i_0})\approx\mu_Y(y_{i_0})\approx1$.

The values of each variable are fuzzified in the same way as X_k on the 10-m depth scale of the borehole in the depth interval [2000; 2500]. In this case, $\mu_X(x)=\min_{k=1,2,3}\mu_{X_k}(x)$, where $X=(X_1, X_2, X_3)$.

$X_{i_0+1}=(X_{1,\,i_0+1}, X_{2,\,i_0+1}, X_{3,\,i_0+1})$, $X_{k,\,i_0+1}$ is the value of the variable $\hat{X}_k$ at point $\hat{h}_{i_0+1}$ for fuzzy regression $\hat{X}_{ik}=\beta_0^{(k)}+\beta_1^{(k)}\tilde{h}_i$ ($k=1,2,3$).

7. Since the fuzzification of the variables X_k($k=1,2,3$) and Y was performed using Equation (25), their left spreads are equal to the right spreads. Therefore, to determine the spreads of the fuzzy variable $Y_i(X_i)(\alpha)$ sufficient for the sample data $\{(l_{X_{i1}}(\alpha), (l_{X_{i2}}(\alpha), (l_{X_{i3}}(\alpha), (l_{Y_i}(\alpha))\ (i=1,\ldots,n;\ n=i_0+1)\}$ we construct the rank regression shown in Equation (22) taking into account the error ρ_1 of forecast y_{i_0+1}. The difference (prediction error on fuzzy regression is calculated (3))

$d(Y_{i_0+1}, \hat{Y}_{i_0+1})$ according to the representation (21) for $\hat{Y}$ in LR-form and formula (27).

Thus, the total (overall) forecast error $\hat{Y}_{i_0+1}$ is calculated as $\varepsilon_1 = d(Y_{i_0+1}, \hat{Y}_{i_0+1}) d(X_{i_0+1}, \hat{X}_{i_0+1}) \rho_1$, where $\hat{X}_{i_0+1}$ is the output variable value $\hat{X}$ and $d(X_i, \hat{X}_i) = \prod_{k=1}^{3} d(X_{ik}, \hat{X}_{ik})$. According to the rolling forecasting scheme, by shifting the set of indices $I_1=\{1, ..., i_0+1\}$ one step forward $\Delta i=0.005$, we obtain the set of indices $I_2=\{2, ..., i_0+2\}$. Then, the forecast error $\hat{Y}_{i_0+2}$, obtained by the fuzzy regression presented in Equation (3) will be equal to $\varepsilon_2 = \varepsilon_1 d(Y_{i_0+2}, \hat{Y}_{i_0+2}) d(X_{i_0+2}, \hat{X}_{i_0+2}) \rho_2$, where ρ_2 is calculated using Equation (30) by substituting x_{n_0} with x_{n_0+1}, etc. At step n of this procedure, a prediction of the value (x_{n_0+n}) is obtained with the following error:

$$\varepsilon_n = \varepsilon_{n-1} d(Y_{i_0+\nu}, \hat{Y}_{i_0+\nu}) d(X_{i_0+\nu}, \hat{X}_{i_0+\nu}) \rho_n$$

If $\varepsilon_n < \varepsilon_0$ (ε_0=acceptable forecast error) and $(i_0+\nu)\times$ $\times 2000 < 2500$, then the procedure will continue. If $\varepsilon_n \geq \varepsilon_0$, then we should switch to predicting the thickness loss at 100-m intervals [13].

The following predictions are obtained using this algorithm:

$$i = i_0 + 1 = 39 \text{ (2390 m)}, \ \hat{y}_{39} = 0.357;$$
$$i = i_0 + 1 = 40 \text{ (2400 m)}, \ \hat{y}_{40} = 0.362$$

Conclusion

1. Flaw detection measurements based on the analysis of the electromagnetic field generated by a probe located inside the casing or tubing can be significantly expensive, particularly when dealing with deep wells and aggressive environments. Therefore, accurately predicting the results of flaw detection examinations at greater depths of pipe immersion in boreholes is crucial. To address this issue, we developed a new method for predicting the maximum loss of pipe thickness at 10 m intervals after a specific depth of the lower part of the pipe is immersed in the borehole, with the maximum thickness loss expressed as a percentage.
2. In our fuzzy regression model, the main parameters of the formation within the borehole (temperature, density, and viscosity of the hydrocarbon mixture) are used as input variables, while the maximum pipe thickness loss in the designated depth intervals is the output variable. Both input and output variables are preliminarily fuzzified in the normality proposition of the resulting fuzzy variables.
3. The prediction process involves sliding in the direction of increasing losses with a forward depth shift of 10 m. The error of the resulting prediction is evaluated using a numerical method to calculate the difference between fuzzy numbers. The output predictions are made by combining rank fuzzy regression and crisp nonlinear regression. To achieve this, the values of each reservoir parameter are interpolated from a 100-m well depth scale to a 10-m depth scale, i.e., to the values given in table 2. using an interpolation Lagrange polynomial.
4. To predict the maximum thickness loss, the best polynomial regression model of its dependence on the depth of pipe immersion in the borehole is constructed using the data in table 1. This method is used to predict the thickness loss on a 10-m depth scale.
5. The variables are fuzzified into normal fuzzy numbers using an adaptive algorithm to construct LR shape membership function.
6. The proposed approach is demonstrated through flaw detection measurements conducted in a specific borehole within the Bulla oil and gas field in Azerbaijan. This methodology can also be applied to address similar challenges related to damage and deterioration of complex equipment in other industries.

References

1. Koskov, V. N., Koskov, B. V., Yushkov, I. R. (2010). Complex assessment of the condition and operation of oil wells using field geophysical methods. *Perm: Perm State Technical University.*

2. Gusseinov, G. G. (2012). Upgraded technology of accidents liquidation in wells and mechanisms used to achieve the purpose. *Oilfield Engineering*, 8, 36-38.

3. (2019). Technical guidelines for the use of geophysical surveys and operations in oil and gas wells. SOCAR Guidelines. *Baku: SOCAR*.

4. Feyziyeva, G. E. (2022). Results of the use of geophysical methods in assessing the working condition of oil and gas wells. *Equipment.Technologies.Materials*, 9, 38-42.

5. Conover, W. J., Iman, R. L. (1976). On some alternative procedures using ranks for the analysis of experimental designs. *Communications in Statistics - Theory and Methods*, 5(14), 1349-1368.

6. Choi, S. H., Buckley, J. J. (2008). Fuzzy regression using least absolute deviation estimators. *Soft Computing*, 12, 257-263.

7. Iman, R. L., Conover, W. J. (1979). The use of the rank transform in regression. *Technometrics*, 21, 499-509.

8. Cai, K. Y. (1993). Parameter estimations of normal fuzzy variables. *Fuzzy Sets and Systems*, 55, 179-185.

9. Nahmas, S. (1978). Fuzzy variables. *Fuzzy Sets and Systems*, 1, 97-110.

10. Jung, H-E., Yoon, J. H., Choi, S. H. (2014). Fuzzy linear regression using rank transform method. *Fuzzy Sets and Systems*, 274, 97-108.

11. Hong, D. H. (2001). Parameter estimations of mutually T-related fuzzy variables. *Fuzzy Sets and Systems*, 123, 63-71.

12. Bakhvalov, N. S. (1973). Numerical methods (analysis, algebra, ordinary differential equations). *Moscow: Nauka.*

13. Habibov, I. A., Dyshin, O. A., Feyziyeva, G. E. (2023). Application of a rank fuzzy regression model to predict the technical conditions of well pipes. In: *The fifth Eurasian RISK-2023, 18*, 5(75), 600-601