Application Study of Univariate Linear Regression Model in Seepage Pressure Analysis of Earth Dam
Taking Dongzhen Reservoir in Fujian Province as an Example

Hui Tian, Genquan Qin*, Chengwei Liao, Ping Lin

1 Fujian Province Disaster Prevention and Mitigation Information Processing and Applied Enterprise Engineering Technology Application Research Center, Fuzhou 350000, China
2 Strong Technology Co., LTD., Fuzhou 350000, China
* Corresponding author’s e-mail: qingenquan@163.com

Abstract: Based on the typical non-deterministic relationship between the seepage pressure of the earth dam and the upstream water level, this paper proposes a method for establishing a linear regression model for seepage pressure analysis of the earth dam; Taking the Dongzhen Reservoir Dam as an example, two historical data representing osmometers BS3 and BS4 were selected for simulation application, and the reliability test of the regression model was carried out using the correlation coefficient test method. The results showed that the correlation coefficients of the reservoir water level and the seepage pressure monitoring values were 0.911 and 0.850, respectively, and the correlations were good and in accordance with the normal law; By applying the prediction and control functions of the built model, the prediction interval of the monitoring results of the osmometer at different reservoir water levels and the safe variation range of the reservoir water level at different seepage pressure values are obtained. Subsequently, the model was used to analyze and verify the data of all 29 piezometers in Dongzhen Reservoir, and the attribution analysis of the relevant differences.

1. Introduction
Leakage has always been a problem threatening the safety of the dam that are in place for water storage. Based on statistics collected from home and abroad, accidents caused by leakage accounted for up to 30%-40% of all accidents for embankment dams [1]. Therefore, scholars have been doing a series of qualitative analysis and engineering practices by studying outliers found in dam seepage, data variation, safety monitoring system and model predictions to evaluate the safety and trends of dam projects [2-4]. For example, Lv Gaofeng used finite element to calculate the seepage field before and after outlier increases and compared the values to study possible influences of outlier increases on clay inclined wall and seepage stability, offering references for future operation of projects [5]. Xiao Shiyuan carried out a two-dimensional finite element seepage analysis on the typical section of the dam through data-based parameter inversion to monitor the reservoir dam, and concluded that the permeability of the dam filling material had a crucial impact on the safe operation of the project [6]. Based on the analysis of seepage pressure variation, the prediction model for seepage pressure is also being built up. Qin Jihui compared the stepwise regression model (SRA), wavelet neural network (WNN), and the model based on stepwise regression and wavelet neural network (SRA-WNN), and concluded that the former two models had more errors than the SRA-WNN model in the prediction of earth dam seepage pressure while the input
factors of SRA-WNN model is more practical as indicators of dam seepage [7]. Liu Cong took a project as an example to study the relationship between seepage and stress coupling using the iterative coupling algorithm, where the correlation between soil permeability coefficient and soil volume strain functioned as the coupling bridge [8]. He concluded that the coupling effect of seepage field and stress field in engineering would improve the safety of the embankment dam, which could be utilized in engineering design. Lei Yan established a partitioned model of finite element seepage based on characteristics of specific projects, and carried out the stable and unstable seepage analysis on the embankment dam of the reservoir to determine if the project had seepage failure, providing theoretical references for similar projects [9]. Targeting the warning system of seepage, Wang Yu used rescaled range analysis (R/S analysis) to explain fractal scale invariance of the seepage time series, revealing that the seepage time series was characterized by tendency and randomness [10].

Among the existing researches, there are few studies on the law of seepage and solutions to problems in this regard. Nor is there report of model introduction to predict and evaluate seepage of dam projects. The paper, based on seepage data obtained from Dongzhen Reservoir, established a unitary linear regression model based on statistical regression analysis of the non-deterministic relationship between reservoir level and seepage pressure, which was applied in predicting seepage pressure and controlling water level. The paper also conducted an overall analysis of seepage relevance at Dongzhen Reservoir based on data collected from 29 seepage pressure monitoring instruments in addition to evaluate the reliability and applicability of the model.

2. Deduction of Seepage Regression Model of Embankment Dam

2.1. Data pre-processing

Previous studies have shown that outliers are inevitable in dam data, which have a significant impact on regression analysis and the estimation of parameters, thus causing a large residual and affecting the coupling of regression equations and the results of prediction [11]. Therefore, it is crucial to examine outliers found in dam data as the first step of data analysis. The commonly used statistical test for outliers include Laida Guidelines (3σ), Grubbs criterion, Dixon and T-test (3S, also known as the Romanovsky criterion), which are based on normal distribution and small probability [12]. Based on characteristics of the data, we adopted Grubbs.

Grubbs criterion: Statistical analysis is practical with the presence of outliers in the sequence of normal distribution.

\[
G = \frac{|x_d - \bar{x}|}{S_n} > \alpha , \text{ where } S_n = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}
\]

where, \(x_j\) is the sequence monitoring value; \(x_d\) is the monitoring value to be tested; \(\bar{x}\) is the average value of the sequence monitoring value; \(S_n\) is the standard deviation of the sequence monitoring value; and \(\alpha\) represents the significance level under Grubbs criterion [13].

2.2. Establishing a data model

There are multiple methods and models for dam seepage monitoring and prediction, including statistical regression model, time series model, support vector machine model and neural network model [7], among which the statistical regression model is theoretically optimal due to its simple structure, fewer involvement of parameters and efficiency in processing non-linear and high-dimensional problems. The seepage coefficient, seepage prevention and drainage of the dam body were taken as the fixed boundary conditions in the paper while the water level of the reservoir was the only variable that could change the seepage pressure of the dam, based on which the unitary linear regression equation of reservoir water level and dam seepage pressure was established.
2.2.1. Unitary linear regression model. Assume that there are two variables \( x \) and \( y \) with certain but unknown correlation, then the unitary linear regression model takes the form of:

\[
\begin{align*}
  y &= a + bx + \varepsilon \\
  \varepsilon &\sim N(0, \sigma^2)
\end{align*}
\]  

(2)

among which \( a, b \) and \( \sigma^2 \) are unknown parameters unrelated to \( x \); \( \varepsilon \) is a random and unpredictable variable. Assume \( \mathbb{E}\varepsilon = 0 \), and \( \mathbb{D}\varepsilon = \sigma^2 \). As \( \varepsilon \) is a random variable, then \( y \) is also a random variable. Therefore, we have:

\[
y \sim N(a + bx, \sigma^2)
\]  

(3)

To make interval estimation and hypothesis test for parameters, a set of data series with sample size of \( n \) is assumed as: \( (x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n) \). The estimated values of \( a \) and \( b \) obtained by regression analysis are \( \hat{a}, \hat{b} \) respectively, and based on the theoretical regression equation, we have:

\[
\hat{y} = \hat{a} + \hat{b}x
\]  

(4)

where \( y \) is the regression function of \( x \), and \( \hat{a}, \hat{b} \) are the regression coefficients.

2.2.2. Parameter estimation by least square method. Independent tests were conducted on a subgroup of \( x \) consisting of completely different values \( x_1, x_2, x_3, \ldots, x_n \) to produce corresponding monitoring values of random variable \( y \): \( y_1, y_2, y_3, \ldots, y_n \). Therefore, the unitary regression model can be expressed as:

\[
\begin{align*}
  y_i &= a + bx_i + \varepsilon_i \\
  \varepsilon_i &\sim N(0, \sigma^2)
\end{align*}
\]  

(5)

\[
Q(a,b) = \sum_{i=1}^{n} (y_i - (a + bx_i))^2
\]  

(6)

Select appropriate estimation values of \( a \) and \( b \): \( \hat{a}, \hat{b} \), then \( Q(\hat{a}, \hat{b}) = \min_{a,b} Q(a,b) \)  

(7)

Calculate partial derivatives of \( Q(a,b) \) respectively, and set it at 0, then we have:

\[
\begin{align*}
  \frac{\partial Q}{\partial a} &= -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\
  \frac{\partial Q}{\partial b} &= -2\sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0
\end{align*}
\]  

(8)

Based on sample data \( (x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n) \), we have:

\[
\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i, \bar{x}y = \frac{1}{n} \sum_{i=1}^{n} x_i y_i
\]  

(9)

Put (9) into (8), we have:

\[
\begin{align*}
  a + bx = \bar{y} \\
  ax + bx^2 = \bar{xy}
\end{align*}
\]  

(10)

Sort out the above equations, we get the least squares estimation of \( a \) and \( b \):

\[
\begin{align*}
  \hat{a} &= \bar{y} - \hat{b} \bar{x} \\
  \hat{b} &= \frac{\bar{xy} - \bar{x}\bar{y}}{\bar{x}^2 - (\bar{x})^2}
\end{align*}
\]  

(11)
The regression equation of y with respect to x:

\[ \hat{y} = \hat{a} + \hat{b}x = \bar{y} + \hat{b}(x - \bar{x}) \]  

(12)

The graph of the equation is called a regression line, which goes through the point \((x, y)\) with a slope of \(\hat{b}\). Put \(x_i\) into (12), we have:

\[ \hat{y}_i = \bar{y} + \hat{b}(x_i - \bar{x}) \]  

(13)

where \(\hat{y}_i\) is called the regression value.

2.2.3. Significance test of the regression model. For variables \(x\) and \(y\), whether there is a correlation or not, the sample regression line is always available, though it needs to be tested and verified. According to the statistical theory, the significance test of the regression coefficient in equation (4) can be conducted to test the correlation significance of \(x\) and \(y\), that is, to test the significance of the regression model. Three methods, including F-test, T-test and Correlation Coefficient \((R)\), are recommended [14]. The paper adopted Correlation Coefficient \((R)\) for this purpose, and we have:

\[ R = \frac{\bar{xy} - \bar{x}\bar{y}}{\sqrt{\bar{x}^2 - (\bar{x})^2} \sqrt{\bar{y}^2 - (\bar{y})^2}} \]  

(14)

The correlation coefficient \(R\) shows the degree of linear correlation between \(x\) and \(y\). According to the length of data sample \(n\) and the given significance level \(\alpha\) (usually 0.05 or 0.01), the critical value \(c\) can be obtained by referring to the critical value table. When \(|R| > c\), it means that there is a linear correlation between \(x\) and \(y\). Otherwise, there is no linear correlation between the two.

3. Prediction and Control of the Regression Model

3.1. Prediction

The model is used to predict the seepage pressure value and its variation range when a series of determined values of reservoir water level is fixed. After determining the empirical regression equation, for given \(x = x_0\), put \(x_0\) into the empirical regression equation, and we have:

\[ \hat{y}_0 = \hat{a} + \hat{b}x_0 \]  

(15)

where \(\hat{y}_0\) is the predicted value of \(y_0\).

In the unitary regression model, assume \(y_1, y_2, y_3, \ldots, y_n\) are independent from each other, then we have:

\[ y_0 = a + bx_0 + \epsilon_0, \epsilon_0 \sim N(0, \sigma^2) \]  

and the regression value \(\hat{y} = \hat{a} + \hat{b}x\). Based on normal distribution, we have:

\[ y_0 - \hat{y}_0 \sim N(0, \left[ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n(x^2 - (\bar{x})^2)} \right] \sigma^2) \]  

(16)

For the confidence coefficient \(1 - \alpha\), from \(P = \left[ |\hat{y} - \bar{y}| < t_{\alpha/2}(n-2) \right] = 1 - \alpha\)

(17)

We have the expected \(1 - \alpha\) confidence interval of \((\hat{y}_0 - \delta(x_0), \hat{y}_0 + \delta(x_0))\)

(18)

where \(\delta(x_0) = t_{\alpha/2}(n-2) \sqrt{\frac{S_E}{n-2}} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{n(x^2 - (\bar{x})^2)}}\), \(S_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\). The approach is adopted by the following calculations.
3.2. Control

Control is the inverse problem of prediction. When the value of seepage pressure $y$ is evaluated in the interval of $(y_1,y_2)$, and $x_1 < x < x_2$, the monitoring value $y$ with respect to $x$ needs to fall in the $1-\alpha$ confidence interval of $(y_1,y_2)$ to calculate water level $x_1$ and $x_2$, which are the solutions of the following equations:

\[
\begin{align*}
\hat{a} + \hat{b}x - \delta(x) & \geq y_1 \\
\hat{a} + \hat{b}x + \delta(x) & \leq y_2
\end{align*}
\]

(19)

Convert the equations, we get $x_1$ and $x_2$:

\[
\begin{align*}
x_1 &= (y_1 - \hat{a} + \hat{b} \cdot u_{\alpha/2}) / \hat{b} \\
x_2 &= (y_2 - \hat{a} - \hat{b} \cdot u_{\alpha/2}) / \hat{b}
\end{align*}
\]

(20)

where the simplified $\delta(x) = \hat{\sigma} \cdot u_{\alpha/2}$ is adopted, $\hat{\sigma} = \sqrt{\frac{S_e}{n-2}}$; assume the seepage pressure value $y$ and set $y = y_1 = y_2$, which could be used to solve $x_1$ and $x_2$ from the above equations. If $x_1 > x_2$, the control interval is $(x_2, x_1)$; if $x_2 > x_1$, the control interval is $(x_1, x_2)$.

4. Study on Engineering Application

4.1. Project Introduction

Dongzhen Reservoir [15], located in the middle reaches of Yanshou River of urban Putian, Fujian Province, is a large (Grade 2) water conservancy project with integrated functions of irrigation, flood control, power generation, shipping, breeding and sightseeing. The total storage capacity of the reservoir is 435 million m$^3$. The project consists of a hub and channels. Among them, the hub project involves three types of buildings: dam, water culvert and spillway. The dam is made of clay-core wall, with the highest dam at the height of 58.6 m and 360 m long. When the dam was strengthened in 2014, instruments were put in place to monitor deformation, seepage and seepage pressure, covering seepages of the dam body and around. There are 4 sections in the dam under seepage monitoring and a group of sections around the dam on the left and right bank, with a total of 29 holes of pressure measuring pipes. The general information of the seepage monitoring project is shown in Table 1. The project started on April 25, 2018, using automatic monitoring of seepage pressure at the frequency of half an hour/time in the initial stage and 8 hours/time during normal operation.

| Section | Shaft Distance | Gauge No. | Site/Height (m) | Section | Shaft Distance | Gauge No. | Site/Height (m) |
|---------|----------------|-----------|-----------------|---------|----------------|-----------|-----------------|
| B0+090  | Z0-001         | BS1       | Core Wall /54.32|         | Z0+009         | BS16      | Foundation /44.38|
|         |                | BS2       | Foundation /49.47|         | Z0+071         | BS17      | Foundation /26.70|
|         | Z0+009         | BS3       | Core Wall /47.35|         | Z0-001         | BS18      | Core Wall /61.88 |
|         | Z0+009         | BS4       | Foundation /45.46|         | Z0-001         | BS19      | Foundation /57.30|
| B0+090  | Z0+071         | BS5       | Foundation /31.45|         | B0+310         | Z0+009    | Core Wall /60.72 |
|         | Z0+130         | BS6       | Foundation /32.44|         | B0+036         | Z0-001    | Foundation /62.78 |
| B0+155  | Z0-001         | BS7       | Core Wall /53.19|         |                 | RR1       |                 |
4.2. Data pre-processing

20 sets of monitoring data from the seepage pressure gauge BS3 and BS4 in the section of B0 + 090 of Dongzhen Reservoir were included in the study. Based on the above formula, the monitoring data collected from the seepage pressure gauge were preprocessed and analyzed as the first step. Grubbs criterion was used to test the monitoring values of two seepage pressure gauges and the water level of the reservoir. The G statistics and critical values (significance level $\alpha = 0.05$) were obtained, as shown in Table 2. The results showed that the G statistic (4.25) of the water level at 8:00 on December 21, 2018 was only greater than the critical value (2.56). The test found that the corresponding water level of the value was 35.3m, lower than the dead water level of the reservoir (50m), which was obviously outlier and should be excluded from the data.

Table 2. Seepage Pressure Values from BS3 and BS4 and G statistics

| Date         | Water level Data BS3 | Water level Data BS4 | Water level Data Reservoir | G Statistics BS3 | G Statistics BS4 | G Statistics Reservoir | Critical Value $\alpha = 0.05$ |
|--------------|----------------------|----------------------|---------------------------|------------------|------------------|-------------------------|-----------------------------|
| 2018/12/22 0:00 | 72.88               | 72.65               | 75.69                     | 1.54             | 1.41             | 0.21                    | 2.56                        |
| 2018/12/21 16:00 | 72.88               | 72.64               | 75.7                      | 1.54             | 1.65             | 0.21                    | 2.56                        |
| 2018/12/21 8:00 | 72.89               | 72.68               | 35.3                      | 1.27             | 0.67             | 4.25                    | 2.56                        |
| 2018/12/22 0:00 | 72.88               | 72.65               | 75.69                     | 1.54             | 1.41             | 0.21                    | 2.56                        |
| 2018/12/21 16:00 | 72.88               | 72.64               | 75.7                      | 1.54             | 1.65             | 0.21                    | 2.56                        |
| 2018/12/21 8:00 | 72.89               | 72.68               | 35.3                      | 1.27             | 0.67             | 4.25                    | 2.56                        |
| 2018/12/20 16:00 | 72.89               | 72.66               | 75.73                     | 1.27             | 1.16             | 0.21                    | 2.56                        |
| 2018/12/20 8:00 | 72.90               | 72.67               | 75.74                     | 0.99             | 0.92             | 0.21                    | 2.56                        |
| 2018/12/19 8:00 | 72.92               | 72.69               | 75.77                     | 0.44             | 0.43             | 0.22                    | 2.56                        |
| 2018/12/19 0:00 | 72.92               | 72.69               | 75.78                     | 0.44             | 0.43             | 0.22                    | 2.56                        |
| 2018/12/18 16:00 | 72.92               | 72.68               | 75.79                     | 0.44             | 0.67             | 0.22                    | 2.56                        |
| 2018/12/18 0:00 | 72.93               | 72.71               | 75.81                     | 0.17             | 0.06             | 0.22                    | 2.56                        |
| 2018/12/17 16:00 | 72.94               | 72.7                | 75.82                     | 0.11             | 0.18             | 0.22                    | 2.56                        |
| 2018/12/17 0:00 | 72.94               | 72.72               | 75.67                     | 0.11             | 0.31             | 0.21                    | 2.56                        |
| 2018/12/16 16:00 | 72.94               | 72.71               | 75.85                     | 0.11             | 0.06             | 0.23                    | 2.56                        |
| 2018/12/16 0:00 | 72.95               | 72.72               | 75.87                     | 0.39             | 0.31             | 0.23                    | 2.56                        |
| 2018/12/16 0:00 | 72.95               | 72.72               | 75.87                     | 0.39             | 0.31             | 0.23                    | 2.56                        |
| 2018/12/15 16:00 | 72.88               | 72.70               | 75.88                     | 1.54             | 0.18             | 0.23                    | 2.56                        |
| 2018/12/15 0:00 | 72.96               | 72.73               | 75.9                      | 0.66             | 0.55             | 0.23                    | 2.56                        |
| 2018/12/14 0:00 | 72.98               | 72.75               | 75.92                     | 1.21             | 1.04             | 0.23                    | 2.56                        |
| 2018/12/13 16:00 | 72.98               | 72.74               | 75.93                     | 1.21             | 0.79             | 0.24                    | 2.56                        |
4.3. Establishing the regression model

According to the pre-processed data and statistical theory, unitary linear regression models concerning monitoring values from two seepage pressure gauges and the water level of the reservoir were established respectively, which were tested for significance based on equation (14). The results are shown in Table 3 and Figure 1.

From Table 3, the correlation coefficients of both the BS3 and BS4 seepage pressure gauges are above 0.80, higher than the critical value of 0.444, indicating that the monitoring values of the two seepage pressure gauges are significantly correlated with the water level of the reservoir. Moreover, the regression values obtained by the regression model basically conform to the monitoring data (Figure 1), indicating that the unitary linear regression model is effective and practical for prediction and control. The regression model shows that the monitoring values of the seepage pressure corresponds to varying water levels of the reservoir, with changes in the slope less than 1(0.3500 and 0.3677 respectively), indicating a stable relationship between seepage pressure of the dam and the water level of the reservoir when there is no drastic changes in seepage.

Table 3. Results of Regression Model

| Gauge | Unitary Linear Regression Model | Correlation Coefficient | Critical Value | Significance |
|-------|---------------------------------|-------------------------|---------------|-------------|
| BS3   | $y=0.3500x+46.4028$             | 0.911                   | 0.444         | Significant |
| BS4   | $y=0.3677x+44.8316$             | 0.850                   | 0.444         | Significant |

Figure 1. Comparison and Analysis of Seepage Pressure Values and Regress Values from BS3 and BS4

4.4. Prediction and Control

According to the unitary linear regression model, when the water level of the reservoir and the range of seepage pressure values are given, the corresponding predicted values and range of seepage pressure and the range of water level are presented as shown in Table 4.

The paper set the water level of the reservoir at 75.78 and used the established regression model to predict and analyze the monitoring values of two seepage pressure gauges. The predicted seepage pressure values were 72.922 and 72.693 respectively, both within the rage, indicating the credibility of the prediction. The predicted ranges of the seepage pressure were (75.803, 75.916) and (75.645, 75.953) respectively, also within the range. Based on the control theory, the paper set the value of seepage pressure for BS3 and BS4 at 72.95 and 72.70 respectively, both the $1 - \alpha$ confidence intervals of which were within the range. The predicted ranges of water level under controlled analysis were (75.803, 75.916) and (75.645, 75.953) respectively, all within the range of normal operation, indicating a stable correlation between the seepage pressure value and the water level of the reservoir under stable seepage conditions in the regression model.
Table 4. Prediction and Control of the Regression Model

| Gauge       | Unitary Linear Regression Model | Prediction | Control |
|-------------|---------------------------------|------------|---------|
| BS3         | \(y=0.350x+46.4028\)            | Given water level of reservoir 75.78 | Predicted value of seepage pressure 72.922 |
| BS4         | \(y=0.3677x+44.8316\)           | Predicted range of seepage pressure (72.888,72.955) | Given seepage pressure value 72.95 |

4.5. Application of the model

Based on data collected from the unitary regression model that monitored data of 29 seepage pressure gauges in Dongzhen Reservoir from September 1, 2019 to November 19, 2019, a regression equation was established to analyze outliers in the monitoring series (see Table 5). According to the results of the correlation analysis (see Table 6), there are two gauges, namely BS17 and RL1 whose correlation coefficient of the regression equation was below the critical value \((n=100, \alpha=0.05)\). The other gauges were highly correlated with the water level of the reservoir in the linear equation, with the coefficient up to 0.981 (BS16). The correlation between the seepage pressure and the water level of reservoir is affected by various factors such as site location, geological conditions and dam-building materials. In terms of monitoring site, the BS17 osmotic pressure gauge is the dam foot of the backwater slope of the dam, far away from the upstream water-retaining boundary of the reservoir. RL1 is located at the right and left dam abutment. Affected by the groundwater level of the mountain on both sides or the distance of the seepage path, the data of the seepage pressure gauge and the water level of the upstream may show insignificant correlation, which was partially proved by the correlation of the seepage pressure gauges in Dongzhen Reservoir in the regression equation.

Table 5. Summary Results of Regression Model of Seepage Pressure Monitoring Series

| Gauge | Regression Equation | Correlation Coefficient | Critical Value | Significance     |
|-------|---------------------|--------------------------|----------------|-----------------|
| BS1   | \(y = 0.5025x+31.1798\) | 0.891                    | 0.197          | Significant     |
| BS2   | \(y = 1.0066x+5.3355\) | 0.904                    | 0.197          | Significance Error |
| BS3   | \(y = 0.2972x+49.5646\) | 0.350                    | 0.197          | Significant     |
| BS4   | \(y = 0.3938x+42.1538\) | 0.436                    | 0.197          | Significant     |
| BS5   | \(y = -0.2916x+58.0253\) | -0.547                   | 0.197          | Significant     |
| BS6   | \(y = 0.0691x+30.3704\) | 0.353                    | 0.197          | Significant     |
| BS7   | \(y = 0.1936x+59.0868\) | 0.921                    | 0.197          | Significant     |
| BS8   | \(y = 0.2217x+56.1950\) | 0.938                    | 0.197          | Significant     |
| BS9   | \(y = 0.4725x+34.3961\) | 0.922                    | 0.197          | Significant     |
| BS10  | \(y = 1.2399x-27.6590\) | 0.873                    | 0.197          | Significant     |
| BS11  | \(y = -0.2520x+51.4600\) | -0.571                   | 0.197          | Significant     |
| BS12  | \(y = -0.2618x+51.6497\) | -0.475                   | 0.197          | Significant     |
| BS13  | \(y = 0.8474x+11.1718\) | 0.952                    | 0.197          | Significant     |
| BS14  | \(y = 0.4045x+42.7325\) | 0.979                    | 0.197          | Significant     |
| BS15  | \(y = 0.4056x+44.3026\) | 0.981                    | 0.197          | Significant     |
| BS16  | \(y = 0.031x+38.359\) | 0.025                    | 0.197          | Insignificant   |
| BS17  | \(y = 0.6635x+25.2948\) | 0.865                    | 0.197          | Significant     |
| BS18  | \(y = 0.6933x+23.2099\) | 0.869                    | 0.197          | Significant     |
| BS19  | \(y = 0.3935x+45.7808\) | 0.828                    | 0.197          | Significant     |
Table 6. Outlier Analysis in Seepage Pressure Monitoring Series

| Gauge | Correlation Coefficient | Analysis |
|-------|--------------------------|----------|
| BS2   | 0.904                    | The overall seepage pressure values were greater than the water level of the reservoir, indicating errors in the cardinal numbers of the gauge |
| BS5   | -0.547                   | Simultaneous drop of water level of the seepage pressure gauge and the reservoir, while the dropping rate of the gauge slower than the reservoir |
| BS11  | -0.571                   | Simultaneous drop of water level of the seepage pressure gauge and the reservoir, while the dropping rate of the gauge slower than the reservoir |
| BS12  | -0.475                   | The gauge is located in the intersection between the backward slope of the dam and the dam foundation, far from the axis of the dam. |
| BS17  | 0.865                    | The overall seepage pressure values were greater than the water level of the reservoir, indicating errors in the cardinal numbers of the gauge |
| RR1   | -0.719                   | Mainly influenced by the mountain groundwater confluence on the right side |
| RR2   | -0.293                   | Mainly influenced by the mountain groundwater confluence on the right side |
| RL1   | -0.101                   | Mainly influenced by the mountain groundwater confluence on the left side |

5. Conclusion

(1) Based on the significant yet undetermined correlation between the water level of the reservoir and the seepage pressure of the dam body, the paper used historical data of seepage pressure to predict and control water level for the safe operation of the reservoir through the unitary regression model using the least squares parameter estimation.

(2) The unitary linear regression model based on seepage pressure data from Dongzhen Reservoir shows good correlation on the whole, indicating the practical values of this approach in analyzing seepage pressure of embankment dam. According to the distribution and seepage of the seepage field, it is concluded that the location of the buried seepage gauge is a significant factor influencing the correlation between the seepage pressure and the reservoir. The results of the regression equation based on data from 29 seepage pressure gauges of Dongzhen Reservoir verified the theory to some extent.

(3) Drastic changes of the water level would prevent the formation of stable seepage in the dam body, making it impractical to deduct the regression model by the unitary linear equation based on the water level and seepage pressure of the dam. Thus, multivariate linear regression analysis is needed by taking time, material and other factors into account.

(4) Considering the complexity of the seepage boundary and the stages involved in the seepage, the fixed-boundary is recommended for multivariate analysis of the seepage pressure. The correlation between factors influencing seepage of the dam and their part of contribution need to be studied and integrated into the matrix of the multivariate linear regression analysis.
References

[1] ZHENG Hui. (2011) Research and Implementation of Dam Seepage Monitoring System Based on GPRS. Beijing Jiaotong University, Beijing.

[2] HE Jinping. (2010) Theory and application of dam safety monitoring. China Power Press, Beijing.

[3] SHA Chenggang. (2014) The Earth-Rock Dam Seepage and Stability Analysis Based on the Geo-Studio. Lanzhou University of Technology, Lanzhou.

[4] LIU Jie, TAN Shaohua, LI Jianlin, etc. (2005) Analysis of Seepage Pressure Based on Improved Genetic Algorithm. Hydropower Automation and Dam Monitoring, 29(3): 74-76.

[5] LU Gaofeng, WANG Yue, ZHOU Jianbo, etc. (2017) Analysis on abnormal measured value by osmometer and seepage stability in clay sloping core dam. Dam and safety, 1: 47-50.

[6] XIAO Shiyan, LIU Xuexiang, YANG Xingyu, etc. (2019) Dam Seepage Safety Evaluation Based on Safety Monitoring Data Inversion. Journal of Yellow River Conservancy Technical Institute, 31 (4): 1-5.

[7] QIN Jihui, WU Yunxiong, GU Yanchang. (2018) A seepage pressure prediction model based on the stepwise regression and wavelet neural network for the embankment dam. Journal of Safety and Environment, 18(5):1670-1674.

[8] LIU Cong. (2013) Research on Seepage-Stress Coupling of Earth-rock Dams and Its Application. Xihua University, Chengdu.

[9] LEI Yan, WANG Zhiyang, ZHANG Haiyang. (2019) Analysis and Evaluation of Stable-unsteady Seepage Flow of Earth-rock Dam in Chijinxiia Reservoir. Shaanxi Water Resources,10:7-8.

[10] WANG Yu, CHEN Chao, CHEN Xiangyong. (2018) Study on prediction and early-warning of seepage safety for earth-rockfill dam. Water Resources and Hydropower Engineering, 49(11): 68-74.

[11] WU Zhongru. (2003) Safety Monitoring Theory and Its Application of Hydraulic Structures. Higher Education Press, Beijing.

[12] WEI Xiaoyi. (2019) Exploration of gross error processing method of dam automatic monitoring data. Theoretical Research, 227.

[13] ZHOU Zhonghan, WANG Hanjiang, LI Mei, etc. (2007) Using DPS to remove outliers in measurement data. Metrology Technology, 61-63.

[14] SHENG Su, XIE Shiqian, PAN Chengyi, etc. (2010) Probability Theory and Mathematical Statistics (Fourth Edition). Higher Education Press, Beijing.

[15] LIN Jinliang. (2018) Application of construction technology for anti-seepage treatment of Dongzhenn reservoir water conservancy project. Heilongjiang Science and Technology of Water Conservancy, 8 (42): 179-180.