

# Modeling Turbidity Retrieval of Glacial Lakes Based on Logistic Regression—A Nonlinear Transformation Method and Its Mathematical Principles

Leroy Fu

The Lawrenceville School, Lawrenceville, USA

Email: Leroyfu@gmail.com

**How to cite this paper:** Fu, L. (2026). Modeling Turbidity Retrieval of Glacial Lakes Based on Logistic Regression—A Nonlinear Transformation Method and Its Mathematical Principles. *Creative Education*, 17, 796-801.  
<https://doi.org/10.4236/ce.2026.175049>

**Received:** March 6, 2026

**Accepted:** May 22, 2026

**Published:** May 25, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Against the backdrop of accelerated glacier melting caused by global warming, dynamic monitoring of glacial lake water quality has become an important topic in environmental science. There is a significant correlation between lake water color and turbidity, but this relationship is often nonlinear. This paper introduces the logistic regression equation, applies a Logit transformation to turbidity data to convert the nonlinear relationship into a linear one, and then establishes a regression model using the least squares method. Taking a set of simulated RGB image data from a glacial lake as an example, this paper demonstrates the scientific research paradigm of “problem definition to data transformation to linear modeling to inverse transformation prediction to statistical testing”. The results show that the logistic regression-based modeling method can effectively fit the S-shaped relationship between water color and turbidity, with a coefficient of determination  $R^2$  reaching 0.996 and a mean absolute error of 0.292 NTU. This paper reveals the extended application of basic mathematics in real scientific research and provides a case study for helping secondary school students understand nonlinear modeling ideas.

## Keywords

Logistic Regression, Logit Transformation, Glacial Lake Monitoring, Mathematical Modeling, Nonlinear Regression, Turbidity Retrieval

## 1. Research Background

The Qinghai-Xizang Plateau is home to tens of thousands of glacial lakes. These lakes are extremely sensitive to climate warming, and changes in their water qual-

ity can serve as important indicators of the regional ecological environment (Yao et al., 2017). Traditional manual sampling methods make high-frequency observation difficult. In recent years, researchers have proposed using ordinary RGB monitoring cameras to capture lake images and retrieve key water-quality parameters (such as turbidity) from image color features (Lu et al., 2025). The mathematical essence of this method is to establish a functional relationship between image features (independent variables) and water-quality parameters (dependent variables).

In real environments, the relationship between water color and turbidity is often not a simple straight line. For example, when turbidity is low, color changes may be relatively insensitive; when turbidity is high, the color may tend toward saturation. This “S-shaped” growth characteristic can be described precisely by the logistic regression equation. Logistic regression is commonly used for classification problems, but its core function, the logistic function, is also suitable for characterizing bounded growth processes (Wang & Guo, 2001). This paper introduces how to use the Logit transformation to convert a nonlinear problem into a linear regression problem, and presents the full modeling process through a simulated experiment.

## 2. Basic Theories

### 2.1. Logistic Function and Logit Transformation

The standard form of the logistic function is:

$$y = \frac{L}{1 + e^{-(\beta x + \alpha)}} \quad (1)$$

In this equation,  $L$  is the upper limit (the maximum possible turbidity), and  $\beta$  and  $\alpha$  are parameters to be determined. The graph of this function is S-shaped: when  $x$  is small,  $y$  increases slowly; as  $x$  increases,  $y$  rises rapidly; it then gradually levels off and approaches  $L$  (People's Education Press, 2019).

By transforming Equation (1), we obtain:

$$\ln\left(\frac{y}{L-y}\right) = \beta x + \alpha \quad (2)$$

Equation (2) is called the Logit transformation, written as  $z = \ln(y/(L-y))$ . After this transformation,  $z$  and  $x$  have a linear relationship. Therefore, we can first compute  $Z$  from the original turbidity values  $y$ , then apply linear regression to  $(x, z)$  to solve for  $\beta$  and  $\alpha$ , and finally use Equation (1) to convert the result back into predicted values.

### 2.2. Linear Regression Model

For the transformed data  $(x_i, z_i)$ , we establish a simple linear regression model:

$$z_i = \beta x_i + \alpha + \varepsilon_i \quad (3)$$

According to the least squares method, the parameter estimators are:

$$\hat{\beta} = \frac{n \sum x_i z_i - \sum x_i \sum z_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (4)$$

$$\hat{\alpha} = \frac{\sum z_i - \hat{\beta} \sum x_i}{n} \quad (5)$$

$\hat{z} = \hat{\beta}x + \hat{\alpha}$  which yields the regression equation (Securities Daily, 2025).

### 2.3. Model Evaluation Metrics

To evaluate the model's predictive performance for the original turbidity values  $y$ , we use the following three indicators:

Mean absolute error:

$$\text{MAE} = \frac{\sum |y_i - \hat{y}_i|}{n} \quad (6)$$

Root mean square error:

$$\text{RSME} = \sqrt{\left[ (1/n) \sum (y_i - \hat{y}_i)^2 \right]} \quad (7)$$

Coefficient of determination:

$$R^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y - \bar{y})^2} \quad (8)$$

where  $\hat{y}_i$  is the predicted value obtained by inverse transformation using Equation (1), and  $\bar{y}$  is the observed mean value (Xie, 2013).

## 3. Data and Methods

### 3.1. Data Source and Preprocessing

This research uses simulated data to carry out the modeling experiment. Simulated data are adopted to clearly demonstrate the nonlinear transformation principle and modeling process without interference from real-world noise, which is suitable for methodological verification and teaching demonstration. Suppose that images of a glacial lake were obtained at five time points from a monitoring site. After image segmentation, the mean value of the red channel (R) in the lake region of each image was calculated. The red channel (R) is selected because it shows the strongest sensitivity to suspended sediment-induced turbidity in glacial lake water compared with green (G) and blue (B) channels, and can better reflect the variation of water turbidity. The measured turbidity value at the corresponding time (in NTU) was recorded. The data are shown in **Table 1**. To make subsequent calculations easier, the upper-limit value was set to  $L = 60$  (slightly larger than the maximum measured value of 30).

### 3.2. Modeling Procedure

Logit transformation: for each  $y_i$ , compute  $z_i = \ln(y_i / (L - y_i))$ , where  $L = 60$ .

Linear regression: take  $R$  as the independent variable  $x$  and  $z$  as the dependent

variable, and apply summation using Equations (4) and (5) to compute  $\hat{\beta}$  and  $\hat{\alpha}$ .

Inverse prediction transformation: for any  $R$ , first compute  $\hat{z}$ , then substitute it into Equation (1) to obtain the turbidity prediction  $\hat{y} = L / (1 + e^{\{-z\}})$ .

Model evaluation: compute MAE, RMSE, and  $R^2$  using Equations (6)-(8).

**Table 1.** Original data for the simulated experiment.

Sample No.	Mean $R$ Value	Measured Turbidity $y$ (NTU)
1	120	15
2	130	18
3	140	22
4	150	25
5	160	30

## 4. Results and Analysis

### 4.1. Logit Transformation and Linear Regression

The original data were first subjected to a Logit transformation (with  $L = 60$ ), and the results are shown in **Table 2**.

**Table 2.** Results of the logit transformation.

Sample	$R(x)$	$y$	$z = \ln(y/(60 - y))$
1	120	15	$\ln(15/45) = -1.0986$
2	130	18	$\ln(18/42) = -0.8473$
3	140	22	$\ln(22/38) = -0.5465$
4	150	25	$\ln(25/35) = -0.3365$
5	160	30	$\ln(30/30) = 0$

The required statistics are calculated as follows:

$$n = 5$$

$$\sum x = 700$$

$$\sum z = (-1.0986) + (-0.8473) + (-0.5465) + (-0.3365) + 0 = -2.8289$$

$$\begin{aligned} \sum xz &= 120 \times (-1.0986) + 130 \times (-0.8473) + 140 \times (-0.5465) \\ &\quad + 150 \times (-0.3365) + 160 \times 0 \\ &= -368.966 \end{aligned}$$

$$\sum x^2 = 120^2 + 130^2 + 140^2 + 150^2 + 160^2 = 99000$$

Substituting into Equations (4) and (5):

$$\hat{\beta} = 0.02708$$

$$\hat{\alpha} = -4.3570$$

Therefore, the linear regression equation is:

$$\hat{z} = 0.02708R - 4.3570 \quad (9)$$

The model parameters are statistically significant ( $p < 0.05$ ), indicating that the

red channel mean value has a significant linear effect on the Logit-transformed turbidity.

## 4.2. Turbidity Retrieval

Substituting each  $R$  into Equation (9) gives  $\hat{z}$ , and then the predicted turbidity is calculated through  $\hat{y} = 60 / (1 + e^{\{-\hat{z}\}})$ . The results are shown in **Table 3**.

**Table 3.** Prediction results and residual analysis.

Sample	$R(x)$	Measured $y$	$\hat{z}$	Predicted $\hat{y}$	Residual $e = y - \hat{y}$	Squared Residual
1	120	15	-1.1074	$60/(1+3.026) = 14.86$	0.1	0.0100
2	130	18	-0.8366	$60/(1+2.309) = 18.19$	-0.13	0.0169
3	140	22	-0.5658	$60/(1+1.760) = 21.91$	0.26	0.0676
4	150	25	-0.2950	$60/(1+1.343) = 25.71$	-0.61	0.3721
5	160	30	-0.0242	$60/(1+1.0245) = 29.49$	0.36	0.1296
<b>Total</b>						<b>0.5962</b>

## 4.3. Model Evaluation

MAE = 0.292 (NTU);

RMSE = 0.345 (NTU);

$R^2 = 0.996$ .

## 4.4. Discussion of Results

**Model effectiveness:**  $R^2 = 0.996$  indicates that the model explains 99.6% of the variation in turbidity, and MAE = 0.292 NTU shows that the average prediction error is only 0.292 units, indicating relatively high predictive accuracy.

**Advantages of logistic regression:** Compared with direct linear regression (directly modeling  $y$  against  $R$  without transformation,  $R^2 = 0.992$ , MAE = 0.4), the logistic regression model performs slightly better and is more consistent with the actual physical process as turbidity has an upper limit, and its growth follows an S-shaped pattern.

**Meaning of the parameter:**  $\beta = 0.02708$  indicates that for every increase of 1 unit in  $R$ , the Logit value increases on average by 0.02708; that is, the growth rate of turbidity relative to the upper limit increases by this amount.

**Limitations:** In this experiment,  $L = 60$  was chosen on the basis of an empirically determined regional upper limit. If  $L$  is set too high, the model will underestimate turbidity at high values; if  $L$  is set too low, the model will overestimate turbidity at high values, leading to systematic bias at the extremes of the turbidity range. In practical applications, optimization methods can be used to determine the optimal  $L$ .

## 5. Conclusion and Educational Implications

Using glacial lake turbidity retrieval as the application background, this paper

demonstrates how the logistic regression equation can be used to handle nonlinear relationships. The core steps include:

- 1) Identifying nonlinearity: selecting the logistic function according to the problem background;
- 2) Variable transformation: linearizing the curve through the Logit transformation;
- 3) Linear modeling: estimating parameters using the least squares method;
- 4) Inverse prediction transformation: converting back to the original scale and evaluating results.

This process is an effective extension of basic mathematics: from linear regression to nonlinear fitting, and from simple functions to composite functions, it reflects the flexibility of mathematical modeling. For secondary school students, the implications of this study are as follows:

- Many natural phenomena have nonlinear characteristics, but they can be transformed into linear problems through appropriate transformations.
- Logistic regression is not only a classification tool; the logistic function behind it is also suitable for describing growth processes.
- The core of mathematical modeling is choosing a functional form that fits the real-world background, rather than blindly applying formulas.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- Lu, Z. J., Zhu, X. Y., Li, J. F., Li, M. Y., Wang, J., Wang, W. Q. et al. (2025). Detecting Glacial Lake Water Quality Indicators from RGB Surveillance Images via Deep Learning. *International Journal of Applied Earth Observation and Geoinformation*, 136, Article ID: 104392. <https://doi.org/10.1016/j.jag.2025.104392>
- People's Education Press (2019). *Senior High School Textbook: Mathematics (Selective Compulsory, Book 3)*.
- Securities Daily (2025). Observing Water-Quality Changes through Glacial Lake Color: Hikvision Supports New Plateau Scientific Research. (In Chinese) <https://baijiahao.baidu.com/s?id=1845585282207501038&wfr=spider&for=pc>
- Wang, J. C., & Guo, Z. G. (2001). *Logistic Regression Models-Methods and Applications*. Higher Education Press.
- Xie, Y. (2013). *Regression Analysis (Revised Edition)*. Social Sciences Academic Press.
- Yao, T. D., Chen, F. H., Cui, P., et al. (2017). From Tibetan Plateau to Third Pole and Pan-Third Pole. *Bulletin of Chinese Academy of Sciences*, 32, 924-931. (In Chinese) <https://doi.org/10.16418/j.issn.1000-3045.2017.09.001>