**Jurnal Varian**
Vol. 8, No. 1, October 2024, pp 67–78
e-ISSN: 2581-2017

# Weibull Regression Model Analysis of Mahakam River Water Pollution Potential

**Zalva Pradipa, Suyitno, Meiliyani Siringoringo**
Universitas Mulawarman, Samarinda, Indonesia

**Article Info**

**Article history:**

Received : 12-22-2023
Revised : 09-16-2024
Accepted : 10-09-2024

**Keywords:**

DO;
Mahakam River;
Pollution Potential;
Weibull Regression.

**ABSTRACT**

Mahakam River has a vital role in the lives of the people of the East Kalimantan province, including providing a raw source of clean water. The multi-activity of the Mahakam River watershed, as a water traffic lane, mining, fisheries, hotels, restaurants, and resident houses, has the potential to produce waste into the water. Increasing waste in the water flow can increase the pollution potential of river water, threatening people's health. Therefore, precaution is necessary. In this research, statistical prevention was proposed, providing information to the East Kalimantan people regarding the factors affecting the pollution potential of the Mahakam River through Weibull regression (WR) modeling on dissolved oxygen (DO) data 2022. Research data was secondary data provided by the Life Environmental Department of East Kalimantan province. The WR model is a Weibull distribution that is directly influenced by covariates. WR model consists of Weibull survival regression, cumulative distribution regression, hazard regression, and Weibull mean regression. This research aims to obtain the factors affecting the pollution potential and to provide the pollution potential information of Mahakam River 2022. The research concluded that factors influencing the pollution potential of the Mahakam River were watercolor degree and nitrate concentration. Applying the WR model to DO data 2022 was able to provide the pollution potential information of Mahakam River, namely the probability of river water isn't polluted is 0.6555, or the probability of the polluted river water is 0.3445, the pollution rate is 6 locations are polluted for every 10 mg/L DO, and the DO average of river water is 5.7450 mg/L. Increasing water color degree and nitrate concentration will decrease the probability of the Mahakam River being polluted, increase the probability of the Mahakam River being polluted, increase the pollution rate, and reduce the DO of Mahakam River water.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v8i1.3699

**Corresponding Author:**

Suyitno,
Statistics Study Program, Faculty of Mathematics and Natural Sciences,
Universitas Mulawarman, Indonesia,
Email: suyitno.stat.unmul@gmail.com



How to Cite:
Pradipa, Z., Suyitno, S., & Siringoringo, M. (2024). Weibull Regression Model Analysis of Mahakam River Water Pollution Potential. *Jurnal Varian*, *8*(1), 67-78.


## A. INTRODUCTION

The Weibull distribution is one of the continuous random variable distributions that is widely used to analyze time data (Fajriati et al., 2022). Discussion of the Weibull distribution is still limited to parameter estimation and data distribution testing, so developing a distribution model directly influenced by covariates is necessary. Furthermore, the Weibull distribution that is directly affected by covariates is called the Weibull Regression (WR) model (Rinne, 2008). WR model consists of Weibull survival regression, Weibull cumulative distribution regression, Weibull hazard regression model, and Weibull regression model for the mean (Suyitno et al., 2022). Many problems in the medical, health, and environmental fields are solved through WR modeling (Collett, 2023; Inayah et al., 2021).

The WR model is mostly applied to time data, such as research discussed by Azizah et al. (2023) which discusses the application of the WR model to the hospitalization time of COVID-19 patients. However, data in the environmental field (not time data) is often found to be Weibull-distributed, and the environmental field problems can be solved using WR modeling. In this research, an application of the WR model was developed to non-negative continuous data (not time data), which was applied to Dissolved Oxygen (DO) data of Mahakam River water. Some research related to environmental data, such as research by Olivera & Heard (2019), which discusses increases in extreme rainfall events, and Shu & Jesson (2021), which discusses wind energy. Moreover, Azad et al. (2014) discuss wind energy conversion.

DO is the amount of oxygen dissolved in water. DO is one of the important parameters in determining water quality. Good water quality is determined by high dissolved oxygen content (Yuliantari et al., 2021). DO can come from the photosynthesis process of aquatic plants, where the amount is not fixed depending on the number of plants, and from the atmosphere (air) that enters the water at a limited speed Fardiaz (1992). Based on Government Regulation Number 22 of 2021 concerning the implementation of environmental protection and management, the minimum limit of DO for class 1 water quality is 6 mg/L (Pemerintah Republik Indonesia, 2022). Based on this regulation, the DO value can be classified into two categories: DO content more than or equal to 6 mg/L indicates good water quality, and DO content less than 6 mg/L indicates polluted water quality. Some factors that are thought to affect DO are watercolor degree, nitrate concentration, detergent, water pH, total phosphate, Total Dissolved Solids (TDS), ammonia concentration, and ferrum.

Mahakam River has a vital role in the lives of East Kalimantan people. Many activities along the Mahakam watershed, such as industrial, agricultural, forestry, mining, and residential sectors, have the potential to produce domestic and non-domestic waste (Suyitno et al., 2022). The amount of waste produced by these activities could cause the Mahakam River to be polluted and threaten public health. Therefore, this study proposes to apply the WR model to DO data to statistically prevent Mahakam River water pollution. This model can provide information regarding the pollution potential of Mahakam River water and the factors influencing it. This information is very useful to the East Kalimantan people as it contributes to preventing pollution in the Mahakam River.

Through WR modeling, the factors influencing water pollution potential can be identified, and information on the pollution potential of the Mahakam River can be obtained. The obtained pollution potential information consists of the probability that the Mahakam River water is not polluted, the probability that the Mahakam River water is polluted, the pollution rate, and the DO average. This modeling also discusses the interpretation of the WR model to DO data. Interpreting the WR model to DO data yields insights crucial for developing effective strategies to protect the Mahakam River and ensure its sustainability for future generations. Based on this description, this research will discuss the analysis of the potential for Mahakam River water pollution using WR modeling with a case study of DO water pollution indicator data in 2022. The parameter estimation method was maximum likelihood estimation (MLE), and the calculation uses R opensource software.

## B. RESEARCH METHOD

### 1. Data

The research variables consist of response variables and independent variables (covariates) presented in Table 1.

Table 1. Research Variables

| Variable Name | Notation | Variable Type | Data Type |
|---|---|---|---|
| DO | $Y$ | Response Variable | Continuous |
| Degree of Color | $X_1$ | Covariates | Continuous |
| Detergent | $X_2$ | Covariates | Continuous |
| Nitrate Concentration | $X_3$ | Covariates | Continuous |
| pH | $X_4$ | Covariates | Continuous |
| Total Phosphate | $X_5$ | Covariates | Continuous |
| TDS | $X_6$ | Covariates | Continuous |
| Ammonia Concentration | $X_7$ | Covariates | Continuous |
| Ferrum | $X_8$ | Covariates | Continuous |

This research used secondary data from the East Kalimantan Provincial Environmental Office. The population of this study is along the Mahakam River. The sample points on the Mahakam River flow from upstream to downstream Mahakam, which have been determined by the East Kalimantan Provincial Environmental Office in 2022. The research procedure carried out is given in Figure 1.

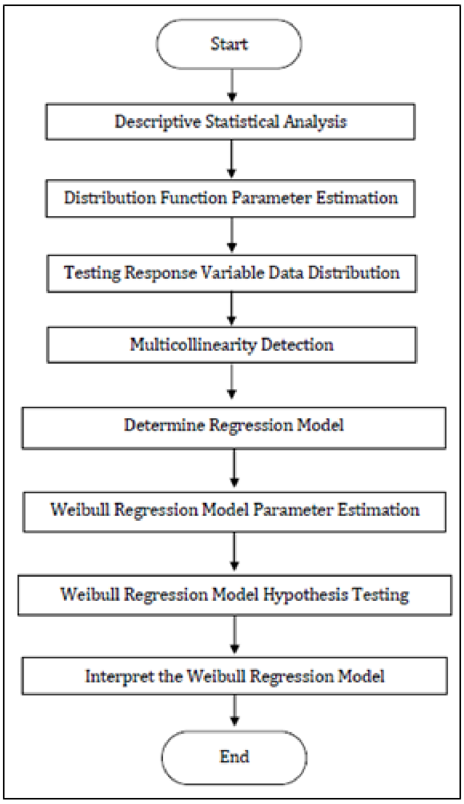



Figure 1. Analysis Step Flow Chart

## 2. Estimation of Weibull Distribution Parameters

The probability density function (PDF) of the random variable $Y$ (Rinne, 2008), which has a scaled-shape version of the Weibull distribution given by

$$f\left(y\right)=\frac{\gamma}{\eta}\left(\frac{y}{\eta}\right)^{\gamma-1}\exp\left[-\left(\frac{y}{\eta}\right)^{\gamma}\right], \tag{1}$$

where $\eta$ is the scale parameter, and $\gamma$ is the shape parameter. The PDF given by Equation (1) can be written in another form as

$$f\left(y\right)=\lambda\gamma y^{\gamma-1}\exp\left[-\lambda y^{\gamma}\right] \tag{2}$$

with $\lambda=\eta^{-\gamma}$ where $\lambda$ is the new scale parameter. The cumulative distribution function and the survival function, respectively, are given by,

$$F\left(y\right)=1-\exp\left[-\lambda y^{\gamma}\right], \tag{3}$$

and

$$S\left(y\right)=\exp\left[-\lambda y^{\gamma}\right]. \tag{4}$$

The Weibull distribution hazard function is given by,

$$h\left(y\right)=\lambda\gamma y^{\gamma-1}. \tag{5}$$

Based on Equation (1), the r-th moment and the mean (expectation) of the random variable $Y$ respectively is


DOI: https://doi.org/10.30812/varian.v8i1.3699

$$E\left(Y^r\right)=\int_0^{\infty} y^r f\left(y\right) dy=\lambda^{-\frac{r}{\gamma}}\,\Gamma\left(\frac{r}{\gamma}+1\right), \tag{6}$$

and

$$\mu_Y=E\left(Y\right)=\lambda^{-\frac{1}{\gamma}}\,\Gamma\left(\frac{1}{\gamma}+1\right). \tag{7}$$

One method for estimating Weibull distribution parameters is Maximum Likelihood Estimation (MLE). The likelihood function based on PDF given by Equation (1) is defined by

$$L(\boldsymbol{\theta}_0)=\prod_{i=1}^{n}\lambda\gamma y_i^{\gamma-1}\exp\left(-\lambda y_i^{\gamma}\right), \tag{8}$$

The log-likelihood function based on the likelihood function is

$$\ell\left(\theta_0|y\right)=\ln\left[L\left(\theta_0|y\right)\right]=\sum_{i=1}^{n}\ln\lambda+\ln\ \gamma+\left(\gamma-1\right)\ln\ y_i-\lambda\, y_i^{\gamma}. \tag{9}$$

$\hat{\theta}_0$ which maximizes the likelihood function is obtained from the first derivative of the log-likelihood Function (9) for all parameters and is equated to zero, namely

$$\frac{\partial\ell\left(\theta_0|y\right)}{\partial\theta_0}=\mathbf{0}, \tag{10}$$

with $\mathbf{0}$ is a 2-dimensional zero vector, and Equation (10) is called the likelihood equation. The left side of equation (10) is a gradient vector of dimension 2, namely

$$\mathbf{g}_1\left(\theta_0\right)=\frac{\partial\ell(\theta_0|y)}{\partial\theta_0}=\left[\frac{\partial\ell\left(\theta_0\right)}{\partial\lambda}\ \frac{\partial\ell\left(\theta_0\right)}{\partial\gamma}\right]^T. \tag{11}$$

Likelihood Equation (10) is a system of interdependent nonlinear equations, so the exact solution to obtain the maximum likelihood (ML) estimator cannot be found analytically. The method to obtain the ML estimator is the Newton-Raphson iterative method (Suyitno et al., 2022).

## 3. Weibull Regression Model Parameter Estimation

The WR model can be mathematically constructed from the Weibull distribution with the scale parameter ($\lambda$) expressed as a function of the covariates or a function of the regression parameters. The scale parameter ($\lambda$) is a positive real number, so it can be expressed as

$$\lambda\left(\mathbf{x}\right)=\exp\left(-\beta^T\mathbf{x}\right)=\exp\left(-\beta_0-\beta_1X_1-\beta_2X_2-\ldots-\beta_pX_p\right), \tag{12}$$

where $\beta^T=\begin{bmatrix}\beta_0 & \beta_1 & \ldots & \beta_p\end{bmatrix}$ and $\mathbf{x}=[1X_1\ldots X_p]^T$.

Substituting Equation (12) into Equation (4) obtained a general Weibull survival regression model, namely

$$S(y,\mathbf{x})=\exp\left(-y^{\gamma}\exp\left[-\beta^T\mathbf{x}\right]\right). \tag{13}$$

Based on Equations (3) and (12), the Weibull cumulative distribution regression model has an expression,

$$F\left(y,\mathbf{x}\right)=1-\exp\left(-y^{\gamma}\exp\left[-\beta^T\mathbf{x}\right]\right). \tag{14}$$

The Weibull hazard regression model can be constructed from Equations (5) and (12), that is,

$$h\left(y,\mathbf{x}\right)=\gamma y^{\gamma-1}\exp\left[-\beta^T\mathbf{x}\right]. \tag{15}$$

The mean Weibull regression model is obtained by substituting Equation (12) into Equation (7), and it has an expression

$$\mu_y(\mathbf{x}) = \Gamma\left(\frac{1}{\gamma}+1\right)\exp\left[\frac{1}{\gamma}\beta^T\mathbf{x}\right]. \tag{16}$$

Based on Equations (13) and (15), we obtain the probability density function (pdf), which is directly influenced by the regression parameters, namely

$$f(Y,\mathbf{x}) = \gamma y^{\gamma-1}\exp\left[-\beta^T\mathbf{x}\right]\exp\left(-y\exp\left[-\beta^T\mathbf{x}\right]\right). \tag{17}$$

Estimating the parameters of the WR model can use the MLE method. The first step in the MLE method is defining the likelihood function (Fajriati et al., 2022). The likelihood function can be written in the form

$$L(\theta) = \prod_{i=1}^{n}\left(\gamma y_i^{\gamma-1}\exp\left[-\beta^T\mathbf{x_i}\right]\right)^{\delta_i}\exp\left(-y_i^{\gamma}\exp\left[-\beta^T\mathbf{x_i}\right]\right), \tag{18}$$

with $\mathbf{x_i} = \begin{bmatrix} x_{i0} & x_{i1} & x_{i2} & \dots & x_{ip}\end{bmatrix}^T$ ; $x_{i0} = 1$ and $\delta_i$ is the classification status of the i-th individual, defined by

$$\delta_i = \begin{cases} 0, \; if\; y_i \geq y^* \\ 1, \; if\; y_i < y^* \end{cases},$$

where $y^*$ is a known positive real constant. Applying the natural logarithm to Equation (18) can be obtained the log-likelihood function, namely

$$\ell(\theta) = \ln L(\theta) = \sum_{i=1}^{n}\left(\delta_i\left[\ln\gamma + (\gamma-1)\ln y_i - \beta^T\mathbf{x_i}\right] - y_i^{\gamma}\exp\left[-\beta^T\mathbf{x_i}\right]\right). \tag{19}$$

The ML estimator of the WR model is obtained by solving the likelihood equation given by

$$\frac{\partial\ell(\theta)}{\partial\theta} = \mathbf{0}, \tag{20}$$

where $\mathbf{0}$ is the zero vector of dimension $p+2$, and the right side of Equation (20) is the $p+2$ dimensional gradient vector. The general form of the gradient vector based on Equation (20) is

$$\mathbf{g}(\boldsymbol{\theta}) = \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \begin{bmatrix}\frac{\partial\ell(\boldsymbol{\theta})}{\partial\gamma} & \frac{\partial\ell(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}^T}\end{bmatrix}^T \tag{21}$$

The likelihood Equation (20) consists of nonlinear equations, so the exact solution to obtain the exact ML estimator cannot be found analytically. The method for solving Equation (18) to obtain an approximation of the ML estimator is the Newton-Raphson iterative method (Khairunnisa et al., 2023).

## 4. Hypothesis testing of Weibull regression model parameters

The regression parameter significance test will be carried out simultaneously and partially. Simultaneous parameter testing is done to determine the significance of the $\boldsymbol{\beta}$ parameter on the response variable simultaneously or to determine the feasibility of the model (Suyitno, 2017). The simultaneous testing hypothesis is

$H_0:\ \beta_1 = \beta_2 = \ldots = \beta_p = 0$
$H_1:$ at least one $\beta_k \neq 0, k = 1,2,\ldots,p$

The test statistic based on the likelihood ratio method is

$$G = 2\left(\ell\left(\hat{\boldsymbol{\theta}}\right) - \ell\left(\hat{\boldsymbol{\theta}}_1\right)\right) \tag{22}$$

where $\hat{\boldsymbol{\theta}} = \begin{bmatrix}\hat{\gamma} & \hat{\beta}_0 & \hat{\beta}_1 & \dots & \hat{\beta}_p\end{bmatrix}$ is the set of parameters under the complete Weibull regression model (population model), and $\ell\left(\hat{\boldsymbol{\theta}}\right)$ is the maximum value of the log-likelihood function in the Equation (19) (Khairunnisa et al., 2023).

Partial parameter testing was done to determine whether certain variables individually affected the Weibull regression model (Suyitno, 2017). The partial testing hypothesis for certain $k$ with $k = 1,2,\ldots,p$ is


DOI: https://doi.org/10.30812/varian.v8i1.3699

$$H_0 \; : \; \beta_k = 0$$
$$H_1 \; : \; \beta_k \neq 0$$

The test statistic is given by

$$Z_s = \frac{\hat{\beta}_k}{\sqrt{var\left(\hat{\beta}_k\right)}} \tag{23}$$

where $Z_s \sim N(0,1)$ for $n \to \infty$, $var\left(\hat{\beta}_k\right)$ is the p+1th diagonal element of the inverse of the Fisher information matrix that is $\left[\mathbf{I_f}\left(\hat{\theta}\right)\right]^{-1}$. The Fisher information matrix is defined by $\mathbf{I_f}\left(\hat{\theta}\right) = -\mathbf{H}\left(\hat{\theta}\right)$ (Suyitno et al., 2022).

## 5. Analysis Steps

The steps of data analysis in this research are as follows:

1) Perform descriptive statistical analysis.
2) Estimating the pdf parameters of the scale-shape version of the Weibull distribution.
3) Testing the distribution of response variable sample data.
   Testing the suitability of data distribution is used to determine the distribution of survival data. One method of testing data distribution is the Kolmogorov-Smirnov test (Dimitrova et al., 2020).
4) Detect multicollinearity between independent variables.
   Multicollinearity in the regression model causes the variance of the parameter estimator to be large. Multicollinearity cases can be detected using the Variance Inflation Factor (VIF) value. A VIF value greater than 10 indicates a case of multicollinearity (Sugiarto et al., 2021).
5) Determine the best Weibull regression model.
   The best model selection is done to get the simplest and best model. Model selection is based on the smallest Akaike Information Criterion (AIC) value (Rencher & Schaalje, 2008).
6) Estimate Weibull Regression model parameters.
7) Hypothesis testing of Weibull regression model parameters simultaneously.
8) Hypothesis testing of Weibull regression model parameters partially.
9) Interpret the Weibull Regression model.
   Interpretation of the Weibull regression model based on the ratio values of the Weibull survival regression model, Weibull cumulative distribution regression, Weibull hazard regression and Weibull mean regression. Weibull hazard regression ratio of the i-th individual based on covariate $X_k$

$$Rh\left(x_{ik}\right) = \exp\left(-\hat{\beta}_k\right), \; k = 1, 2, \ldots, \; p \tag{24}$$

The values of the survival regression ratio and Weibull cumulative distribution based on continuous covariate $X_k$ are respectively

$$RS\left(x_{ik}\right) = \frac{\hat{S}\left(y_i, \mathbf{x}_i | x_{ik} + 1\right)}{\hat{S}\left(y_i, \mathbf{x}_i | x_{ik}\right)}, \tag{25}$$

and

$$RF\left(x_{ik}\right) = \frac{1 - \hat{S}\left(y_i, \mathbf{x}_i | x_{ik} + 1\right)}{1 - \hat{S}\left(y_i, \mathbf{x}_i | x_{ik}\right)}. \tag{26}$$

The ratio value of mean regression based on continuous covariate $X_k$ is

$$R\mu\left(x_{ik}\right) = \exp\left(\frac{1}{\hat{\gamma}}\hat{\beta}_k\right). \tag{27}$$

## C. RESULT AND DISCUSSION

### 1. Descriptive Statistics

Data description is stated in descriptive statistics consisting of mean value (average), maximum value, minimum value, and standard deviation. Descriptive statistics are displayed in Table 2.

Table 2. Descriptive Statistics

| Variable | Average | Maximum | Minimum | Standard Deviation |
|---|---|---|---|---|
| DO ($Y$) | 5.3636 | 7.0 | 3.6 | 1.2254 |
| Degree of Color ($X_1$) | 86.7878 | 170 | 21 | 46.6394 |
| Detergent ($X_2$) | 0.0162 | 0.1300 | 0.0020 | 0.0201 |
| Nitrate Concentration ($X_3$) | 0.3168 | 1.6.0 | 0.02 | 0.4278 |
| pH ($X_4$) | 7.0769 | 7.830 | 6.350 | 0.3210 |
| Total Phosphate ($X_5$) | 0.0309 | 0.180 | 0.025 | 0.0272 |
| TDS ($X_6$) | 59.7557 | 184 | 12 | 47.0111 |
| Ammonia Concentration ($X_7$) | 0.3445 | 5.00 | 0.01 | 0.9299 |
| Ferrum ($X_8$) | 0.6093 | 4.00 | 0.04 | 0.7585 |

Based on the descriptive statistics in Table 2, the average DO is 5.3636 mg/L, below the standard number threshold of 6 mg/L. This shows that, in general, the Mahakam River water is indicated to be polluted.

### 2. Weibull Distribution Parameter Estimation

Weibull distribution parameter estimation was performed on DO data with classification status $\delta_i = 1$ for DO data less than 6 mg/L. It is assumed that the DO data follows Weibull distribution with PDF is given by Equation (2). The result of the parameter estimation using the MLE method is presented in Table 3.

Table 3. Estimated Weibull Distribution Parameters

| Parameters | Estimated |
|---|---|
| Scale ($\lambda$) | 4.6694 |
| Shape ($\gamma$) | 6.7000 |

Based on the parameter estimates in Table 3, the estimated probability density function based on Equation (2) is

$$\hat{f}(y) = 31.28498\, y^{5.7000} \exp\left[-4.6694\, y^{6.7000}\right] \tag{28}$$

and the estimated cumulative distribution function based on Equation (3) is

$$\hat{F}(y) = 1 - \exp\left[-4.6694\, y^{6.7000}\right]. \tag{29}$$

### 3. Weibull Distribution Testing

Testing the distribution of DO data using the Kolmogorov-Smirnov approach. The cumulative distribution function of the sample is expected to be the same as the population distribution function. The cumulative distribution function of the population is $F(y)$. The distribution testing hypothesis formulation is

$H_0 \ : \ F(y) = \hat{F}(y)$

(DO data is Weibull distributed with the distribution function $\hat{F}(y)$ given by Equation (29))

$H_1 \ : \ F(y) \neq \hat{F}(y)$

(DO data is not Weibull distributed).

The results of the calculation of the D test statistic are presented in Table 4.

Table 4. Weibull Distribution Testing of DO Data

| Test Statistics ($D$) | $D_{(18;0,05)}$ | $P_{value}$ | Decision |
|---|---|---|---|
| 0.2331 | 0.3090 | 0.2818 | $H_0$ accepted |

Based on the calculation of the test statistic presented in Table 4, it was decided to accept $H_0$ at the 0.05 significance level. It is supported by the statistical value of $D = 0.23315 < D_{(18;0,05)} = 0.3090$. The conclusion of the hypothesis test states that the DO data is Weibull-distributed.

### 4. Multicollinearity Detection

The initial step of parameter estimation is multicollinearity detection between covariates. Multicollinearity detection is based on the Variance Inflation Factor (VIF) value. VIF value $> 10$ indicates that there is multicollinearity between covariates. Based on result of computation, it was obtained that every VIF value of covariate displayed on Table 2 was less than 10. Conclusion of detecting of multicollinear is there is no multicollinear between covariates, so that WR modeling in this study can involve 8 covariates which is given by Table 2, namely water color degree, detergent, nitrate concentration, pH, total phosphate, TDS, ammonia concentration, and ferrum.

### 5. Parameter Estimation of the Best Weibull Regression Model

WR model in this study consists of the Weibull survival regression model, Weibull cumulative distribution regression model, Weibull hazard regression model, and Weibull regression model for the mean. The best model is obtained based on the minimum AIC value. After the selection process, the best WR model was a model with 2 covariates: color degree and nitrate concentration. The result of estimating the parameter of the WR model using the MLE method is presented in Table 5.

Table 5. Parameter Estimation of Weibull Regression Model

| **Parameter** | **Estimated** |
|---|---|
| $\gamma$ | 5.7028 |
| $\beta_0$ | 12.5639 |
| $\beta_1$ | $-0.0203$ |
| $\beta_3$ | $-1.6266$ |

Based on the results of estimating the Weibull regression model parameters in Table 5, the Weibull survival regression model is obtained as follows:

$$\hat{S}(y, \mathbf{x}) = \exp\left(-y^{5.7028} \exp\left(-12.5639 + 0.0203X_1 + 1.6266X_3\right)\right). \tag{30}$$

The Weibull cumulative distribution regression model is

$$\hat{F}(y, \mathbf{x}) = 1 - \exp\left(-y^{5.7028} \exp\left(-12.5639 + 0.0203X_1 + 1.6266X_3\right)\right). \tag{31}$$

The Weibull hazard regression model is

$$\hat{h}(y, \mathbf{x}) = 5.7028Y^{4.7028} \exp\left(-12.5639 + 0.0203X_1 + 1.6266X_3\right), \tag{32}$$

and the Weibull regression model for the mean is

$$\hat{\mu}_Y(\mathbf{x}) = 0.9250 \exp\left(2.2031 - 0.0035X_1 - 0.2852X_3\right), \tag{33}$$

The Weibull survival regression model shows the probability that the Mahakam River water is not polluted, the Weibull cumulative distribution regression model shows the probability that the Mahakam River water is polluted, the Weibull hazard regression model shows the pollution rate of Mahakam River water, and the Weibull mean regression model shows the average DO of the Mahakam River water.

### 6. Hypothesis Testing of the Best Weibull Regression Model Parameters Simultaneously

Simultaneous parameter testing aims to determine whether the covariates simultaneously affect the RW model. The hypothesis for simultaneous parameter testing is

$H_0$ : $\beta_1 = \beta_3 = 0$
$H_1$ : At least one$\beta_k \neq 0, k = 1,\ 3$


DOI: https://doi.org/10.30812/varian.v8i1.3699

The test statistic is the G statistic with $G \sim \chi^2_2$. The test decision of WR parameters simultaneously is presented in Table 6.

Table 6. Hypothesis Testing Results of WR Model Parameters Simultaneously

| Test Statistics (G) | $\chi^2_{(0,95;2)}$ | $P_{value}$ | Decision |
|---|---|---|---|
| 21.27537 | 5.9914 | $2.3994 \times 10^{-5}$ | $H_0$ was Rejected |

Based on the results of the test statistic calculations shown in Table 6, it was decided to reject $H_0$ at the significance level $\alpha = 0.05$. So, it can be concluded that the WR model is a fit model or that the color degree and nitrate concentration simultaneously influence the WR model. It is supported by the test statistical value $G = 21.27537 > \chi^2_{(0.95;5)} = 5.9914$ and $P_{value} = 2.3994 \times 10^{-5} < \alpha = 0.05$, which shows that the decision to reject H0 was correct with the probability maximum of type 1 error of 5%.

## 7. Hypothesis Testing of the Best Weibull Regression Model Parameters Partially

Partial parameter testing aims to determine the effect of each covariate of the WR model. The partial parameter testing hypothesis for the $\beta_k$ parameter with $k = 0, 1, 3$ is

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0$$

The test statistic used is the $Z_s$ statistic based on Equation (23) with $Z_s \sim N(0, 1)$. The test results are presented in Table 7.

Table 7. Hypothesis Testing Results of WR Model Parameters Partially

| Variable | Parameter | SE | $\|Z_s\|$ | $P_{value}$ | Decision |
|---|---|---|---|---|---|
| Constant ($X_0$) | $\beta_0$ | 2.2192 | 5.6612 | $1.5026 \times 10^{-8}$ | $H_0$ Rejected |
| Degree of Color ($X_1$) | $\beta_1$ | 0.0056 | 3.5691 | $3.5811 \times 10^{-4}$ | $H_0$ Rejected |
| Nitrate Concentration ($X_3$) | $\beta_3$ | 0.4584 | 3.5482 | $3.8779 \times 10^{-4}$ | $H_0$ Rejected |

Based on the statistical value $|Z_s|$ obtained in Table 7, the variables color degree ($X_1$) and nitrate concentration ($X_3$) individually affect the WR model. It is supported by the value $|Z_s|$ variables $X_1$ and $X_3$, respectively, which are 3.5691 and 3.5482 greater than the critical value $Z_{0.975} = 1.96$ and the $P_{value}$ of each variable respectively is $3.5811 \times 10^{-4}$ and $3.8779 \times 10^{-4}$, which is less than $\alpha = 0.05$. The decision to reject H0 on this hypothesis testing was correct, with a confidence level of 95%.

Based on the WR models given by Equations (30), (31), (32), and (33), information on the potential water pollution on the Mahakam River is obtained and displayed in Table 8. Based on Table 8, the average probability that the Mahakam River water is not polluted is 0.6555, and the average probability that the Mahakam River water is polluted is 0.3444. The average pollution rate of Mahakam River water is 0.6083 locations for every 1 mg/L DO, or equal to 6 locations polluted for every 10 mg/L DO. The average DO for Mahakam River water is 5.7450 mg/L. Based on DO content, in general, Mahakam River water is indicated to be polluted because the DO content is less than 6 mg/L.

Table 8. Hypothesis Testing Results of WR Model Parameters Partially

| Regression Model | Probability Average |
|---|---|
| $\bar{\hat{S}}(y, \boldsymbol{x})$ | 0.6555 |
| $\bar{\hat{F}}(y, \boldsymbol{x})$ | 0.3445 |
| $\bar{\hat{h}}(y, \boldsymbol{x})$ | 0.6083 |
| $\bar{\hat{\mu}}(\boldsymbol{x})$ | 5.7450 |

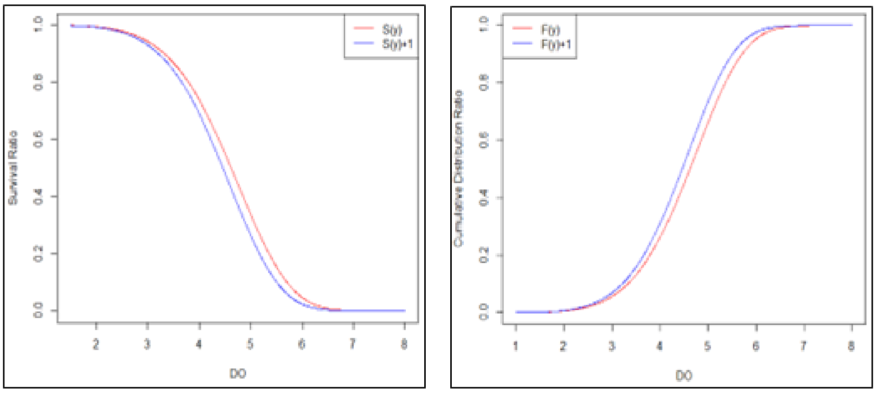
## 8. Interpretation of the Best Weibull Regression Model

The interpretation of the Weibull regression model is based on the ratio value of hazard regression, survival regression, cumulative distribution regression, and mean regression. Interpretation of the Weibull regression model is based on the ratio value of the color degree variable ($X_1$) and nitrate concentration ($X_3$). The calculation results of survival regression ratio, cumulative distribution regression ratio, hazard regression ratio and mean regression ratio can be seen in Table 9.

Table 9. Hypothesis Testing Results of WR Model Parameters Partially

| Regression Model | Weibull Regression Ratio Every 1 Unit Increase of Covariate | |
|---|---|---|
| | $X_1$ | $X_3$ |
| Survival Regression Ratio ($RS(X_k)$) | 0.9965 | 0.5058 |
| Cumulative Distribution Regression Ratio ($RF(X_k)$) | 1.0188 | 3.7227 |
| Hazard Regression Ratio ($Rh(X_k)$) | 1.0205 | 5.0865 |
| Mean Regression Ratio ($R\mu(X_k)$) | 0.9964 | 0.7518 |

Based on Table 9, $RS_1(x_1) = 0.9965$ shows that increasing the color degree will decrease the probability of the Mahakam River not being polluted to 0.9965 times or a decrease of 0.35%. Graphically, decreasing the probability that the Mahakam River water is not polluted after increasing one unit of the color degree can be seen in Figure 2a. Based on Figure 2a, it is known that the blue graph is the graph after increasing 1 unit color degree, while the red graph is the graph before increasing color degree. The blue graph is below the red graph, which shows that the probability of the Mahakam River water not being polluted decreases when the color degree increases. Based on Table 9, $RS(X_3) = 0.5058$ shows that increasing the nitrate concentration will decrease the probability of the Mahakam River not being polluted by 0.5058 times or 49.42%.



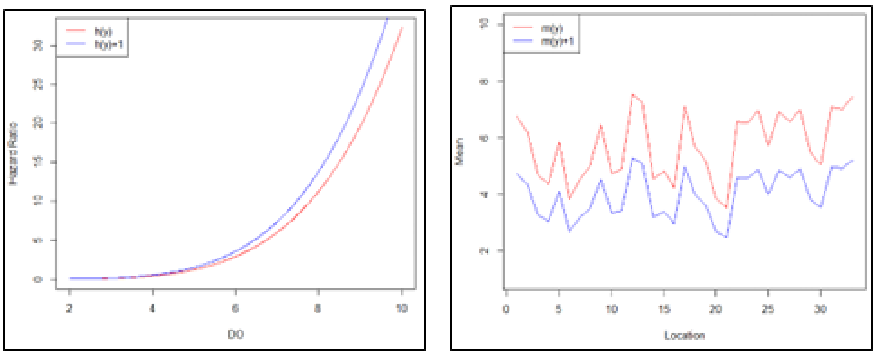

(a) (b)

Figure 2. (a) Probability of Mahakam River Water Not Polluted After Increasing Color Degree (Blue) and Before Increasing Color Degree (Red). (b) The probability of Mahakam River water being polluted after increasing the color degree (blue) and before increasing the color degree (red).

Based on Table 9, $RF_1(x_1) = 1.0188$ shows that increasing the color degree will increase the probability of the Mahakam River being polluted to 1.0188 times or increase by 1.88%. $RF_3(x_3) = 3.7227$ shows that increasing the nitrate concentration will increase the probability of the Mahakam River being polluted to 3.7227 times or increase it by 272.27%. A graph of the increasing probability that the Mahakam River water is polluted after increasing one unit of the color degree ($X_1$) can be seen in Figure 2b. Based on Figure 2b, it is known that the blue graph is the graph after increasing 1 unit color degree, while the red graph is the graph before increasing color degree. The blue graph is above the red graph, which shows that the probability of the Mahakam River water polluted increases when the color degree increases.

As presented in Table 9, $Rh_1(x_1) = 1.0205$ shows that increasing the color degree will increase the pollution rate to 1.0205 times or an increase of 2.05%. $Rh_3(x_3) = 5.0865$ shows that increasing the nitrate concentration will increase the pollution rate to 5.0865 times or increase by 408.65%. A graph of the Mahakam River water pollution rate after increasing one unit of color degree can be seen in Figure 3a. Based on Figure 3a, it is known that the blue graph is the graph after increasing 1 unit color degree, while the red graph is the graph before increasing color degree. The blue graph is above the red graph, which shows that the rate of Mahakam River water pollution increases when the color degree increases.

Table 9 shows $R\mu_1(x_1) = 0.9964$ this means increasing the color degree will decrease the average of DO to 0.9964 times or decrease by 0.36%. $R\mu_3(x_3) = 0.7518$ means that increasing the nitrate concentration will decrease the average of DO to 0.7518 times or decrease by 24.82%. A graph of the decrease in the average of DO Mahakam River water after increasing one unit of the color degree can be seen in Figure 3b.




(a) (b)

Figure 3. (a) Mahakam River Water Pollution Rate After Increasing Color Degree (Blue) and Before Increasing Color Degree (Red). (b) Average DO of Mahakam River Water After the Increasing Color Degree (Blue) and Before the Increasing Color Degree (Red).

Based on Figure 3b, the blue graph shows the Mahakam River water after increasing 1 unit of color degree, while the red graph shows the Mahakam River water before increasing color degree. The blue graph is below the red graph, which means that increasing the color degree will decrease the DO of the Mahakam River water.

Compared with research conducted by Panduwinata et al. (2022), using the Biochemical Oxygen Demand (BOD) indicator, it is known that the factors that influence Mahakam River water pollution are pH, TDS, and water discharge. This study uses DO indicators and obtains factors that influence Mahakam River water pollution: color degree and nitrate concentration. This study also discussed the cumulative distribution regression model, which was not discussed in previous research.

## D. CONCLUSION AND SUGGESTION

Based on the results and discussion carried out in the previous chapter, it is known that factors that affect the DO of Mahakam River water are the degree of color degree and nitrate concentration. Increasing color degree and nitrate concentration, respectively, will decrease the probability that the Mahakam River is not polluted, increase the probability of the Mahakam River being polluted, increase the pollution rate, and decrease the average of DO. Based on the Weibull regression model, it is known that the pollution potential of the Mahakam River water is that the average probability that the Mahakam River water is not polluted is 0.6555, and the average probability that the Mahakam River water is polluted is 0.3445. The average rate of polluted Mahakam River water is 0.6083 locations for every 1 mg/L DO, or equal to 6 locations are polluted for every 10 mg/L DO, and the average DO for Mahakam River water is 5.7450 mg/L. Based on DO content, in general, Mahakam River water is indicated to be polluted because the DO content is less than 6 mg/L. Based on the conclusions, it is hoped that the surrounding community can maintain a clean environment. Besides that, factory, mining and residential activities can control waste production.

## REFERENCES

Azad, A. K., Rasul, M. G., Alam, M. M., Uddin, S. M. A., & Mondal, S. K. (2014). Analysis of Wind Energy Conversion System Using Weibull Distribution. *Procedia Engineering*, *90*, 725–732. https://doi.org/10.1016/j.proeng.2014.11.803

Azizah, N., Suyitno, S., & Hayati, M. N. (2023). Pemodelan Laju Kematian Pasien Covid-19 di RSUD Abdul Wahab Sjahranie Samarinda menggunakan Model Regresi Weibull. *Journal of Mathematics Computations and Statistics*, *6*(1), 17–30. https://doi.org/10.35580/jmathcos.v6i1.36379

Collett, D. (2023, May). *Modelling Survival Data in Medical Research* (4th ed.). Chapman; Hall/CRC. https://doi.org/10.1201/9781003282525

Dimitrova, D. S., Kaishev, V. K., & Tan, S. (2020). Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous. *Journal of Statistical Software*, *95*(10), 1–42. https://doi.org/10.18637/jss.v095.i10


DOI: https://doi.org/10.30812/varian.v8i1.3699

Fajriati, N. A., Suyitno, S., & Wasono, W. (2022). Model Regresi Hazard Rate Weibull Kesembuhan Pasien Rawat Inap Demam Berdarah Dengue (DBD) di RSUD Panglima Sebaya Tanah Grogot. *EKSPONENSIAL*, *13*(1), 35–44. https://doi.org/10.30872/eksponensial.v13i1.878

Fardiaz, S. (1992). *Polusi Air dan Udara* (Cet. 1). Kanisius.

Inayah, U. R., Suyitno, S., & Siringoringo, M. (2021). Upaya Pencegahan Pencemaran Air Sungai Mahakam melalui Pemodelan Geographically Weighted Logistic Regression pada Data BOD. *EKSPONENSIAL*, *12*(1), 17–26. https://doi.org/10.30872/eksponensial.v12i1.755

Khairunnisa, S. F., Suyitno, S., & Mahmuda, S. (2023). Weibull Regression Model on Hospitalization Time Data of COVID-19 Patients at Abdul Wahab Sjahranie Hospital Samarinda. *Jurnal Matematika, Statistika dan Komputasi*, *19*(2), 286–303. https://doi.org/10.20956/j.v19i2.22266

Olivera, S., & Heard, C. (2019). Increases in the extreme rainfall events: Using the Weibull distribution. *Environmetrics*, *30*(4), 1–9. https://doi.org/10.1002/env.2532

Panduwinata, H. D., Suyitno, S., & Huda, M. N. (2022). Model Regresi Weibull Pada Data Kontinu yang Diklasifikasikan. *EKSPONENSIAL*, *13*(2), 123–130. https://doi.org/10.30872/eksponensial.v13i2.1051

Pemerintah Republik Indonesia. (2022). *Peraturan Pemerintah Republik Indonesia Nomor 22 Tahun 2021 Tentang Penyelenggaraan Perlindungan dan Pengelolaan Lingkungan Hidup*.

Rencher, A. C., & Schaalje, G. B. (2008). *Linear Models in Statistics* (2nd ed). Wiley-Interscience.

Rinne, H. (2008, November). *The Weibull Distribution: A Handbook*. Chapman; Hall/CRC.

Shu, Z. R., & Jesson, M. (2021). Estimation of Weibull parameters for wind energy analysis across the UK. *Journal of Renewable and Sustainable Energy*, *13*(2), 1–18. https://doi.org/10.1063/5.0038001

Sugiarto, S., Suyitno, S., & Rizki, N. A. (2021). Model Geographically Weighted Univariat Weibull Regression pada Data Indikator Pencemaran Air Dissolve Oxygen di Daerah Aliran Sungai Mahakam Kalimantan Timur Tahun 2018. *EKSPONENSIAL*, *12*(2), 185–192. https://doi.org/10.30872/eksponensial.v12i2.813

Suyitno, S. (2017). Penaksiran Parameter dan Pengujian Hipotesis Model Regresi Weibull Univariat. *EKSPONENSIAL*, *8*(2), 179–183. https://doi.org/10.30872/eksponensial.v8i2.41

Suyitno, S., Nohe, D. A., Purnamasari, I., Siringoringo, M., Goejantoro, R., & Rahmah, M. N. (2022, December). *Monograf: Pemodelan Regresi Weibull Pada Potensi Pencemaran Sungai Mahakam*. Penerbit Perkumpulan Rumah Cemerlang Indonesia. https://www.rcipress.rcipublisher.org/index.php/rcipress/catalog/book/398

Yuliantari, R. V., Novianto, D., Hartono, M. A., & Widodo, T. R. (2021). Pengukuran Kejenuhan Oksigen Terlarut pada Air menggunakan Dissolved Oxygen Sensor. *Jurnal Fisika Flux: Jurnal Ilmiah Fisika FMIPA Universitas Lambung Mangkurat*, *18*(2), 101–104. https://doi.org/10.20527/flux.v18i2.9997