

Upaya Pencegahan Pencemaran Air Sungai Mahakam melalui Pemodelan *Geographically Weighted Logistic Regression* pada Data BOD

Pollute Prevention of The Mahakam River Water using Geographically Weighted Logistic Regression Model on the BOD Data

Ulfah Resti Inayah¹, Suyitno², Meiliyani Siringoringo³

^{1,2}Laboratorium Statistika Terapan FMIPA Universitas Mulawarman

³Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

E-mail: ulfahrestiinayah12@gmail.com

Abstract

Since the early years, Mahakam River has important roles in supporting human needs in East Kalimantan province. Activities around Mahakam watershed such as restaurants, fishery, and industries were in the potential of generating waste around the flow area. The waste consisted of domestic and nondomestic waste. The waste was a threat to the Mahakam River water quality. Water pollution around the Mahakam River was a threat to public health, and therefore, there's a need for precaution. One of the precautions is to give the public information regarding the factors that influence the chances of polluted water in the Mahakam River increased through logistic regression modeling. One way to detect water pollution is to indicate by using Biochemical Oxygen Demand (BOD). BOD data was suspected spatial, therefore the appropriate statistical modeling is Geographically Weighted Logistic Regression (GWLR). GWLR is a regression model that developed from a logistic regression in which parameter estimation is done locally at every observation location. The purpose of the research is to determine the GWLR model on the BOD data of Mahakam River and to find out the factors that influence water pollution at 27 observation points along with the Mahakam River flow. The parameter estimation method is the Maximum Likelihood Estimation (MLE). The spatial weighting is calculated by using the Adaptive Bisquare weighting function and the optimum bandwidth is determined by using Generalized Cross-Validation (GCV) criteria. Research shows that the closed-form of the Maximum Likelihood estimator can't be obtained analytically and the approximation is obtained by using Newton-Raphson (N-R) iterative method. Based on parameter testing of the GWLR model result, it was concluded that the factors were influences the probability of Mahakam River water were polluted based on the BOD indicator was locally and different in each 27 observation locations. The factors that influence locally were water temperature, acidity, Total Dissolved Solids (TDS), ammonia concentration, and water debit, meanwhile, the factors which influence globally were acidity and TDS.

Keywords: Adaptive Bisquare, BOD, GCV, GWLR, MLE, N-R

Pendahuluan

Menurut Hansen, dkk (2017), Sungai Mahakam sejak dulu hingga saat ini memiliki peranan penting dalam menunjang kebutuhan manusia maupun makhluk hidup di sekitarnya khususnya di Provinsi Kalimantan Timur karena digunakan sebagai fungsi ekosistem perairan, mencuci, mandi, sumber air minum, memasak, tempat penampungan air, sarana transportasi, dan lain-lain. Aktivitas di sekitar aliran sungai seperti industri (pabrik), peternakan, perikanan berpotensi menghasilkan limbah domestik maupun nondomestik. Limbah-limbah yang dihasilkan oleh aktivitas tersebut dapat menjadikan air Sungai Mahakam terancam tercemar, oleh karena itu perlu tindakan pencegahan. Salah satu tindakan pencegahan adalah memberikan informasi kepada masyarakat mengenai faktor-faktor yang berpengaruh terhadap peluang meningkatnya air Sungai Mahakam tercemar melalui pemodelan statistika.

Salah satu cara untuk mendeteksi air sungai tercemar adalah melalui indikator BOD. BOD adalah suatu ukuran yang menyatakan jumlah

oksigen terlarut yang digunakan oleh mikroorganisme (biasanya bakteri) untuk mengurai atau mendekomposisi bahan organik (Umalay & Cuvin, 1998). Data BOD merupakan data kuantitatif yang dapat dikategorikan dan diduga merupakan data spasial. Data spasial adalah adalah jenis data yang mengandung informasi atribut dan informasi lokasi, serta terdapat sifat saling ketergantungan antara lokasi dan data.

Data BOD diduga merupakan data spasial yang memuat heterogenitas spasial. Menurut Fotheringham, dkk (2002), heterogenitas spasial terjadi akibat adanya perbedaan pengaruh variabel prediktor terhadap respon antara satu wilayah dengan wilayah lainnya. Pemodelan regresi pada data yang memuat heterogenitas spasial akan menyebabkan nilai estimator yang berbeda-beda, sehingga pemodelan yang sesuai adalah pemodelan yang bersifat lokal. Model regresi yang melakukan penaksiran parameter secara lokal di setiap lokasi pengamatan adalah model *Geographically Weighted Regression* (GWR).

Penelitian-penelitian sebelumnya tentang pemodelan GWR menunjukkan bahwa metode GWR lebih efektif untuk pemodelan regresi pada data spasial dan lebih baik daripada model regresi global. Penelitian yang dilakukan oleh Suyitno, dkk (2016), yaitu *Parameter Estimation of Geographically Weighted Trivariate Weibull Regression (GWTWR) Model* menyimpulkan bahwa model GWTWR lebih baik daripada model global berdasarkan nilai GCV. Penelitian yang dilakukan oleh Desriwendi, dkk (2015) mengatakan model lokal atau *Geographically Weighted Logistic Regression (GWLR)* lebih baik daripada model regresi logistik global berdasarkan nilai AIC.

Ide yang mendasari model GWR adalah penaksiran parameter yang dilakukan pada setiap lokasi pengamatan dan menggunakan pembobot spasial. Hal tersebut menyebabkan model GWR di setiap lokasi pengamatan mempunyai nilai parameter regresi yang berbeda-beda. Lokasi geografis pengamatan dinyatakan dalam koordinat garis lintang (*latitude*) dan garis bujur (*longitude*). Semakin dekat suatu lokasi maka pengaruhnya akan semakin besar, sehingga diberi nilai pembobot yang lebih besar (Fotheringham, dkk., 2002).

Pemodelan GWR pada data respon dikotomis yang memuat heterogenitas spasial disebut dengan *Geographically Weighted Logistic Regression (GWLR)* (Atkinson, dkk., 2003). Penaksiran parameter model GWLR dilakukan pada setiap lokasi pengamatan dan menggunakan pembobot spasial. Pembobot spasial ditentukan menggunakan fungsi pembobot. Fungsi pembobot merupakan fungsi jarak antar lokasi pengamatan dan tergantung pada *bandwidth* atau parameter penghalus. Setiap lokasi pengamatan bisa memiliki nilai *bandwidth* yang konstan atau berbeda-beda. Nilai *bandwidth* yang berbeda-beda pada setiap lokasi disebut *bandwidth* adaptif. Keunggulan penggunaan *bandwidth* adalah memudahkan penentuan *bandwidth* optimum yang menghasilkan model terbaik. Salah satu fungsi pembobot dengan *bandwidth* adaptif adalah fungsi *Bisquare*. Besar nilai pembobot spasial tergantung pada *bandwidth*, sehingga pemilihan *bandwidth* sangat penting. Salah satu metode untuk pemilihan *bandwidth* optimum, yaitu *Generalized Cross-Validation (GCV)* (Fotheringham, dkk., 2002).

Tujuan penelitian ini adalah untuk memperoleh model GWLR data indikator pencemaran air BOD di daerah aliran Sungai Mahakam, memperoleh faktor-faktor yang berpengaruh signifikan terhadap pencemaran air di daerah aliran Sungai Mahakam, dan memperoleh interpretasi model GWLR data indikator pencemaran air BOD di beberapa daerah aliran Sungai Mahakam.

Model Regresi Logistik Biner Global

Model regresi logistik biner global adalah salah satu model regresi untuk menentukan hubungan antara satu atau lebih variabel prediktor yang bersifat kontinu, kategorik, atau gabungan dari kontinu dan kategorik dengan variabel respon yang bersifat dikotomis (berskala nominal atau ordinal dengan dua kategori) (Agesti, 2002).

Regresi logistik biner global termasuk dalam model linier umum (*Generalized Linear Model*) atau GLM. Misalkan Y adalah variabel acak dikotomis berskala nominal, yaitu bernilai 1 atau 0, dengan $y = 1$ menyatakan hasil percobaan sukses dan $y = 0$ menyatakan hasil percobaan gagal. Variabel Y berdistribusi Bernoulli dengan fungsi densitas

$$P(Y = y) = f(y; \pi) = \pi^y (1 - \pi)^{1-y}; y = 0, 1, \quad (1)$$

dengan π adalah parameter. Berdasarkan fungsi densitas pada persamaan (1), diperoleh $P(y = 1) = \pi$ dan $P(y = 0) = 1 - \pi$. Model umum regresi logistik biner global adalah

$$\pi_i = \frac{\exp(\beta^T \mathbf{x}_i)}{1 + \exp(\beta^T \mathbf{x}_i)} \quad (2)$$

dengan $\beta^T = [\beta_0 \ \beta_1 \ \dots \ \beta_p]$ dan $\mathbf{x}_i = [1 \ x_{i0} \ x_{i1} \ \dots \ x_{ip}]^T$.

Penaksiran parameter model regresi logistik biner global dapat menggunakan metode *Maximum Likelihood Estimation (MLE)*. Berdasarkan persamaan (1) dan (2), fungsi *likelihood* didefinisikan oleh

$$L(\beta) = \left(\prod_{i=1}^n (1 + \exp(\beta^T \mathbf{x}_i))^{-1} \right) \exp \left(\sum_{i=1}^n y_i \beta^T \mathbf{x}_i \right), \quad (3)$$

Penaksir *Maximum Likelihood (ML)* model regresi logistik biner global (2) akan lebih mudah diperoleh dengan memaksimalkan fungsi *log-likelihood*. Penerapan logaritma natural kedua ruas persamaan (3) diperoleh

$$\ell(\beta) = \sum_{i=1}^n y_i \beta^T \mathbf{x}_i - \ln(1 + \exp(\beta^T \mathbf{x}_i)), \quad (4)$$

Penaksir ML model regresi logistik (2) diperoleh dengan menyelesaikan persamaan

$$\frac{\partial \ell(\beta)}{\partial \beta} = \mathbf{0}, \quad (5)$$

dengan ruas kiri dari persamaan (5) dinamakan vektor gradien berdimensi $p+1$. Vektor gradien dapat dinyatakan dalam perkalian matriks, yaitu

$$\mathbf{g}(\beta) = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}), \quad (6)$$

di mana $\boldsymbol{\pi} = [\pi_1 \ \pi_2 \ \dots \ \pi_n]^T$ dengan π_i diberikan oleh persamaan (2), $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]^T$, dan \mathbf{X}

adalah matriks data pengamatan berukuran $n \times (p + 1)$, yaitu

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (7)$$

Persamaan *likelihood* (5) terdiri dari persamaan-persamaan nonlinier, sehingga solusi eksak (*closed-form*) untuk mendapatkan penaksir ML model logistik biner tidak dapat ditemukan secara analitik. Metode alternatif untuk menentukan solusi persamaan *likelihood* (5) adalah menggunakan metode iteratif Newton-Raphson (N-R). Algoritma iteratif N-R memerlukan perhitungan vektor gradien dan matriks Hessian. Matriks Hessian berukuran $(p + 1) \times (p + 1)$ dapat dinyatakan dengan perkalian matriks, yaitu

$$\mathbf{H}(\boldsymbol{\beta}) = -\mathbf{X}^T \mathbf{V} \mathbf{X}, \quad (8)$$

dengan \mathbf{X} diberikan oleh persamaan (7) dan \mathbf{V} merupakan matriks diagonal berukuran $n \times n$ dengan elemen diagonal ke- i adalah $\pi_i(1 - \pi_i)$. Berdasarkan matriks Hessian pada persamaan (8), matriks Informasi Fisher dapat dinyatakan dengan

$$\mathbf{I}(\hat{\boldsymbol{\beta}}) = \mathbf{X}^T \hat{\mathbf{V}} \mathbf{X}, \quad (9)$$

Berdasarkan vektor gradien dan matriks Hessian yang masing-masing diperoleh pada persamaan (6) dan (8), maka iterasi N-R dapat dijalankan untuk memperoleh penaksir parameter $\hat{\boldsymbol{\beta}}$ dengan algoritma sebagai berikut.

$$\hat{\boldsymbol{\beta}}^{(r+1)} = \hat{\boldsymbol{\beta}}^{(r)} - [\mathbf{H}(\hat{\boldsymbol{\beta}}^{(r)})]^{-1} \mathbf{g}(\hat{\boldsymbol{\beta}}^{(r)}). \quad (10)$$

Proses iterasi akan berhenti bila sudah terpenuhi kondisi konvergen, yaitu $\|\hat{\boldsymbol{\beta}}^{(r+1)} - \hat{\boldsymbol{\beta}}^{(r)}\| \leq \varepsilon$ dengan ε adalah bilangan yang cukup kecil (misal 10^{-12}).

Pengujian signifikansi parameter model regresi logistik biner global terdiri dari dua tahap, yaitu pengujian signifikansi parameter secara simultan dan secara parsial. Pengujian pertama dilakukan untuk mengetahui signifikansi parameter regresi terhadap variabel respon secara simultan. Hipotesis pengujian simultan adalah $H_0 : \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$

$H_1 : \text{Minimal terdapat satu } \beta_k \neq 0; k = 1, 2, 3, \dots, p$

Statistik uji diberikan oleh

$$G = 2(\ell(\hat{\Omega}) - \ell(\hat{\omega})), \quad (11)$$

di mana $\ell(\hat{\Omega})$ dan $\ell(\hat{\omega})$ dihitung berdasarkan persamaan (4), yaitu

$$\ell(\hat{\Omega}) = \sum_{i=1}^n (y_i \ln(\hat{\pi}_i) + \ln(1 - y_i)(1 - \hat{\pi}_i)), \quad (12)$$

dan

$$\ell(\hat{\omega}) = (n_1 \ln(n_1) + n_0 \ln(n_0) - n \ln(n)), \quad (13)$$

dengan $G \sim \chi_p^2$. Hipotesis nol ditolak pada taraf signifikansi α jika nilai $G > \chi_{(\alpha; p)}^2$ atau $p\text{-value} < \alpha$.

Pengujian parameter secara parsial menggunakan statistik uji *Wald*. Hasil pengujian secara parsial digunakan untuk mengetahui pengaruh setiap variabel prediktor terhadap variabel respon. Hipotesis uji secara parsial untuk parameter β_k dengan nilai k tertentu ($k = 0, 1, 2, \dots, p$) adalah

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0$$

Statistik uji *Wald* diberikan oleh

$$W = \frac{\hat{\beta}_k}{\sqrt{\text{Var}(\hat{\beta}_k)}}. \quad (14)$$

dengan $W \sim N(0, 1)$ dan $\text{Var}(\hat{\beta}_k)$ merupakan elemen diagonal ke- k dari invers matriks Informasi Fisher $[\mathbf{I}(\hat{\boldsymbol{\beta}})]^{-1} = -[\mathbf{H}(\hat{\boldsymbol{\beta}})]^{-1}$, di mana $\mathbf{I}(\hat{\boldsymbol{\beta}})$ diberikan oleh persamaan (9). Daerah kritis pada pengujian hipotesis menolak H_0 pada taraf signifikansi α apabila nilai $|W| > Z_{\alpha/2}$ atau $p\text{-value} < \alpha$.

Pendeteksian Multikolinearitas

Menurut Gujarati (2003), multikolinearitas adalah terjadinya hubungan linier antara variabel prediktor dalam suatu model regresi linier. Salah satu cara pendeteksian multikolinearitas adalah berdasarkan *Variance Inflation Factor* (VIF). Suatu variabel prediktor dikatakan multikolinearitas, jika nilai VIF variabel tersebut lebih dari 10 (Kutner, dkk., 2005). Nilai VIF dihitung berdasarkan rumus

$$\text{VIF}_k = \frac{1}{1 - R_k^2}. \quad (15)$$

dengan R_k^2 adalah koefisien determinasi saat X_k diregresikan dengan variabel prediktor lainnya.

Pengujian Heterogenitas Spasial

Heterogenitas spasial terjadi akibat adanya perbedaan pengaruh variabel prediktor terhadap respon antara satu wilayah dengan wilayah lainnya. Salah satu metode pengujian heterogenitas spasial adalah metode *Glejser*. Hipotesis pengujian heterogenitas spasial adalah (Gujarati, 2003):

H_0 : $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2$
 (Tidak terdapat heterogenitas spasial)
 H_1 : Minimal terdapat satu $\sigma_i^2 \neq \sigma^2; i = 1, 2, \dots, n$
 (Terdapat heterogenitas spasial)
 Statistik uji diberikan oleh

$$F_3 = \frac{(\hat{\alpha}^T \mathbf{X}^T \mathbf{e} - n\bar{e}^2)/p}{(\mathbf{e}^T \mathbf{e} - \hat{\alpha}^T \mathbf{X}^T \mathbf{e})/(n-p-1)} \quad (16)$$

dengan $\hat{\alpha}$ adalah vektor penaksir parameter model regresi *auxiliary*, \mathbf{e} adalah vektor nilai mutlak regresi logistik biner global berdimensi n dengan elemen $\mathbf{e} = [|\hat{e}_1| \ | \ \hat{e}_2| \ \dots \ | \ \hat{e}_n|]^T$, dan \mathbf{X} diberikan oleh persamaan (7). Statistik uji $F_3 \sim F_{(p, (n-p-1))}$. Kriteria penolakan adalah menolak H_0 jika $F_3 > F_{(\alpha, p, (n-p-1))}$ atau p -value $< \alpha$.

Pembobot Spasial pada Model Geographically Weighted Regression

Penaksiran parameter pada model GWR menggunakan pembobot spasial. Pembobot adalah suatu nilai (ukuran) yang menyatakan seberapa besar pengaruh data di suatu lokasi pengamatan dengan lokasi pengamatan yang lain. Data-data di lokasi pengamatan yang berdekatan diberi bobot yang lebih besar karena akan memberikan pengaruh yang lebih kuat daripada data-data di lokasi pengamatan yang lebih jauh (Chasco, dkk., 2007).

Pembobot spasial dihitung menggunakan fungsi pembobot yang bergantung pada jarak antar lokasi pengamatan. Jarak antara lokasi (u_i, v_i) dengan lokasi (u_j, v_j) disimbolkan dengan d_{ij} dan dihitung dengan menggunakan rumus jarak *Euclidean* sebagai berikut (Chasco, dkk., 2007):

$$d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}, \quad (17)$$

Salah satu fungsi pembobot yang digunakan untuk menghitung pembobot spasial adalah fungsi pembobot *adaptive bisquare* yang diberikan oleh

$$w_{ij} = \begin{cases} \left(1 - \left(\frac{d_{ij}}{b_i}\right)^2\right)^2, & \text{untuk } b_i \geq d_{ij} \\ 0, & \text{untuk } b_i < d_{ij} \end{cases} \quad (18)$$

dengan w_{ij} menyatakan pembobot spasial yang diberikan kepada data pengamatan lokasi ke- j untuk model di lokasi pengamatan ke- i dan b_i adalah *bandwidth* untuk penaksiran model GWR pada lokasi ke- i .

Berdasarkan persamaan (18), besarnya nilai pembobot spasial tergantung pada *bandwidth*, sehingga pemilihan *bandwidth* sangat penting. Salah satu metode untuk menentukan *bandwidth* optimum adalah *Generalized Cross-Validation*

(GCV), di mana *bandwidth* optimum akan diperoleh saat nilai GCV minimum. Nilai GCV diperoleh dengan rumus sebagai berikut (Fotheringham, dkk., 2002):

$$GCV = \frac{n \sum_{i=1}^n [y_i - \hat{y}_i(b_i)]^2}{(n-v)}, \quad (19)$$

dengan $\hat{y}_i(b_i)$ adalah taksiran dari y_i yang menggunakan *bandwidth* (b_i) dan $v = \text{tr}(\mathbf{S})$.

Matriks \mathbf{S} diberikan oleh

$$\mathbf{S} = \begin{bmatrix} \mathbf{x}_1^T (\mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{V}(u_1, v_1) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_1, v_1) \mathbf{V}(u_1, v_1) \\ \mathbf{x}_2^T (\mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{V}(u_2, v_2) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_2, v_2) \mathbf{V}(u_2, v_2) \\ \vdots \\ \mathbf{x}_n^T (\mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{V}(u_n, v_n) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_n, v_n) \mathbf{V}(u_n, v_n) \end{bmatrix} \quad (20)$$

dengan $\mathbf{x}_i^T = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}]$, \mathbf{X} diberikan oleh persamaan (7), $\mathbf{W}(u_i, v_i)$ adalah matriks diagonal pembobot spasial berukuran $n \times n$ untuk penaksiran parameter model di lokasi ke- i , di mana elemen diagonal ke- j adalah w_{ij} , dan $\mathbf{V}(u_i, v_i)$ merupakan matriks diagonal dengan elemen diagonal ke- j adalah $\pi_j(u_i, v_i)(1 - \pi_j(u_i, v_i))$.

Pemodelan Geographically Weighted Logistic Regression

Model *Geographically Weighted Logistic Regression* (GWLR) merupakan pengembangan dari model regresi logistik yang diterapkan pada data spasial. Misal koordinat lokasi seluruh pengamatan diketahui, maka model GWLR pada lokasi pengamatan ke- i adalah

$$\pi(u_i, v_i) = \frac{\exp(\boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_i)}{1 + \exp(\boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_i)}; \quad i = 1, 2, \dots, n, \quad (21)$$

Salah satu metode penaksiran parameter model GWLR adalah *Maximum Likelihood Estimation* (MLE). Misal diberikan n data pengamatan variabel prediktor, yaitu $x_{ik}; k = 1, 2, \dots, p; i = 1, 2, \dots, n$ dan n data pengamatan variabel respon dengan $y_i \sim B(1, \pi(u_i, v_i))$, di mana $\pi(u_i, v_i)$ diberikan oleh persamaan (21) serta mempunyai fungsi kepadatan peluang diberikan oleh persamaan (1). Berdasarkan persamaan (1) dan (21), fungsi *likelihood* dengan pembobot spasial untuk penaksiran parameter GWLR pada lokasi ke- i dengan koordinat (u_i, v_i) adalah

$$L(\boldsymbol{\beta}(u_i, v_i)) = \left(\prod_{j=1}^n (1 + \exp(\boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_j))^{-1} \right) \times \exp \left(\sum_{j=1}^n y_j \boldsymbol{\beta}^T(u_i, v_i) \mathbf{x}_j \right)^{w_{ij}}, \quad (22)$$

Penerapan logaritma natural fungsi *likelihood* pada kedua ruas persamaan (22) menghasilkan fungsi *log-likelihood*, yaitu

$$\ell(\beta(u_i, v_i)) = \sum_{j=1}^n w_{ij} (y_j \beta^T(u_i, v_i) \mathbf{x}_j - \ln(1 + \exp(\beta^T(u_i, v_i) \mathbf{x}_j))), \quad (23)$$

Penaksir ML ($\hat{\beta}(u_i, v_i)$) diperoleh dengan cara menurunkan persamaan (23) terhadap $\beta_k(u_i, v_i)$ dengan $k = 0, 1, 2, \dots, p$ dan disamakan dengan nol, maka diperoleh

$$\frac{\partial \ell(\beta(u_i, v_i))}{\partial \beta_k(u_i, v_i)} = \mathbf{X}^T \mathbf{W}(u_i, v_i) (\mathbf{y} - \boldsymbol{\pi}(u_i, v_i)) = \mathbf{0}. \quad (24)$$

Diketahui persamaan (24) adalah implisit (tidak *closed-form*), oleh karena itu untuk menyelesaikan dan mendapatkan penaksir parameter model GWLR digunakan pendekatan numerik. Salah satu pendekatan numerik yang dapat digunakan adalah metode iterasi N-R. Algoritma iteratif N-R membutuhkan perhitungan vektor gradien dan matriks Hessian. Vektor gradien diberikan oleh

$$\mathbf{g}(\beta(u_i, v_i)) = \mathbf{X}^T \mathbf{W}(u_i, v_i) (\mathbf{y} - \boldsymbol{\pi}(u_i, v_i)), \quad (25)$$

Matriks Hessian adalah matriks turunan parsial orde kedua dari fungsi *log-likelihood* $\ell(\beta(u_i, v_i))$ terhadap semua kombinasi komponen-komponen vektor parameter $\beta(u_i, v_i)$. Matriks Hessian dapat dinyatakan dengan

$$\mathbf{H}(\beta(u_i, v_i)) = -\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{V}(u_i, v_i) \mathbf{X}, \quad (26)$$

Berdasarkan matriks Hessian (26), matriks Informasi Fisher dapat dinyatakan dengan

$$\mathbf{I}(\hat{\beta}(u_i, v_i)) = \mathbf{X}^T \hat{\mathbf{V}}(u_i, v_i) \mathbf{X}, \quad (27)$$

Berdasarkan vektor gradien dan matriks Hessian yang masing-masing diperoleh pada persamaan (25) dan (26), maka iterasi N-R dapat dijalankan untuk memperoleh penaksir parameter $\beta(u_i, v_i)$ dengan algoritma sebagai berikut.

$$\hat{\beta}^{(t+1)}(u_i, v_i) = \hat{\beta}^{(t)}(u_i, v_i) - [\mathbf{H}(\hat{\beta}^{(t)}(u_i, v_i))]^{-1} \mathbf{g}(\hat{\beta}^{(t)}(u_i, v_i)). \quad (28)$$

Proses iterasi akan berhenti bila sudah terpenuhi kondisi konvergen, yaitu $\|\hat{\beta}^{(t+1)}(u_i, v_i) - \hat{\beta}^{(t)}(u_i, v_i)\| \leq \varepsilon$ dengan ε adalah bilangan yang cukup kecil (misal 10^{-12}).

Pengujian Kesesuaian Model Regresi Logistik Biner Global dan Model Geographically Weighted Logistic Regression

Pengujian kesesuaian antara model regresi logistik biner global dan model GWLR bertujuan untuk mengetahui signifikansi dari faktor geografis. Hipotesis pengujiannya adalah sebagai berikut.

$$H_0 : \beta_k(u_i, v_i) = \beta_k; k = 1, 2, \dots, p; i = 1, 2, \dots, n$$

$$H_1 : \text{Minimal terdapat satu } \beta_k(u_i, v_i) \neq \beta_k$$

Statistik uji diberikan oleh

$$F_2 = \frac{D(\hat{\beta})/v_1}{D(\hat{\beta}^*)/v_2}, \quad (29)$$

dengan

$$D(\hat{\beta}) = 2(\ell(\hat{\Omega}) - \ell(\hat{\omega})), \quad (30)$$

$\ell(\hat{\Omega})$ dan $\ell(\hat{\omega})$ masing-masing diberikan oleh persamaan (12) dan (13). $D(\hat{\beta})$ berdistribusi χ_p^2 . $D(\hat{\beta}^*)$ didefinisikan oleh

$$D(\hat{\beta}^*) = 2(\ell(\hat{\Omega}_{\text{GWLRL}}) - \ell(\hat{\omega})), \quad (31)$$

di mana

$$\ell(\hat{\Omega}_{\text{GWLRL}}) = \sum_{i=1}^n (y_i \ln(\hat{\pi}_i(u_i, v_i)) + (1 - y_i) \ln(1 - \hat{\pi}_i(u_i, v_i))). \quad (32)$$

$D(\hat{\beta}^*)$ yang diberikan oleh persamaan (31) berdistribusi χ_{np}^2 . Berdasarkan distribusi dari persamaan (30) dan (31), statistik uji $F_2 \sim F_{(p, np)}$. Kriteria pengujian kesesuaian model adalah menolak H_0 jika $F_2 > F_{(\alpha; (p, np))}$ atau $p\text{-value} < \alpha$ (Fathurahman, dkk., 2016).

Pengujian Parameter Model Geographically Weighted Logistic Regression

Pengujian parameter model GWLR dilakukan secara simultan dan secara parsial. Hipotesis pengujian secara simultan adalah sebagai berikut.

$$H_0 : \beta_1(u_i, v_i) = \beta_2(u_i, v_i) = \dots = \beta_p(u_i, v_i) = 0;$$

$$i = 1, 2, \dots, n$$

$$H_1 : \text{Minimal terdapat satu } \beta_k(u_i, v_i) \neq 0;$$

$$k = 1, 2, \dots, p; i = 1, 2, \dots, n$$

Statistik uji diberikan oleh

$$G_2 = 2 \left(\sum_{i=1}^n (y_i \ln(\hat{\pi}_i(u_i, v_i)) + (1 - y_i) \ln(1 - \hat{\pi}_i(u_i, v_i))) - \sum_{i=1}^n (n_{1i} \ln(n_{1i}) + n_{0i} \ln(n_{0i}) - n \ln(n)) \right), \quad (33)$$

dengan $n_1 = \sum_{i=1}^n y_i$, $n_0 = \sum_{i=1}^n (1 - y_i)$, dan n adalah jumlah pengamatan keseluruhan. Statistik uji $G_2 \sim \chi_v^2$, dengan $v = \text{tr}(\mathbf{S})$, di mana \mathbf{S} adalah matriks yang diberikan oleh persamaan (20). Kriteria pengujian adalah menolak H_0 jika nilai $G_2 > \chi_{(\alpha, v)}^2$ atau $p\text{-value} < \alpha$.

Pengujian parameter secara parsial digunakan untuk mengetahui parameter yang berpengaruh secara signifikan terhadap model. Hipotesis uji secara parsial untuk parameter β_k dengan nilai k tertentu ($k = 0, 1, 2, \dots, p$) dan i tertentu ($i = 1, 2, \dots, n$) adalah

$$H_0 : \beta_k(u_i, v_i) = 0$$

(Variabel prediktor X_k tidak berpengaruh terhadap probabilitas air Sungai Mahakam tercemar)

$$H_1 : \beta_k(u_i, v_i) \neq 0$$

(Variabel prediktor X_k berpengaruh terhadap probabilitas air Sungai Mahakam tercemar)

Statistik uji adalah statistik *Wald* diberikan oleh

$$Z_h = \frac{\hat{\beta}_k(u_i, v_i)}{\sqrt{\text{Var}(\hat{\beta}_k(u_i, v_i))}} \quad (34)$$

dengan $Z_h \sim N(0,1)$ dan $\text{Var}(\hat{\beta}_k(u_i, v_i))$ adalah elemen diagonal ke- k dari invers matriks Informasi Fisher pada lokasi (u_i, v_i) yang diberikan oleh persamaan (27). Kriteria pengujian adalah menolak H_0 jika nilai $|Z_h| > Z_{\alpha/2}$ atau $p\text{-value} < \alpha$.

Ukuran Kebaikan Model Regresi Logistik Biner Global dan Model Geographically Weighted Logistic Regression

Ukuran kebaikan model regresi logistik biner global maupun GWLR selain GCV yang diberikan oleh persamaan (19) adalah koefisien determinasi. Koefisien determinasi dihitung menggunakan Pseudo R^2 atau R^2 McFadden's (R_{MF}^2), yaitu

$$R_{MF}^2 = 1 - \frac{\ell(\hat{\Omega}_{\text{GWLR}})}{\ell(\hat{\omega}_{\text{GWLR}})} \quad (35)$$

Suatu model dikatakan lebih baik apabila memiliki nilai GCV terkecil dan nilai R_{MF}^2 paling besar jika dibandingkan dengan model lainnya.

Biochemical Oxygen Demand

BOD adalah suatu ukuran yang menyatakan jumlah oksigen terlarut yang digunakan oleh mikroorganisme untuk mengurai atau mendekomposisi bahan organik (Umalay & Cuvin, 1998). Kandungan BOD berdasarkan standar baku mutu air kelas 1 dapat ditunjukkan pada Tabel 1.

Konsentrasi BOD	Indikasi
≤ 2 mg/l	Tidak tercemar
> 2 mg/l	Tercemar

Sumber: Peraturan Pemerintah Nomor 82 Tahun 2001

Faktor-Faktor yang Memengaruhi Indikator BOD

Suhu air sangat berpengaruh terhadap jumlah oksigen terlarut di dalam air. Peningkatan suhu perairan mengakibatkan meningkatnya konsumsi oksigen oleh mikroorganisme sekitar 2-3 kali lipat, sehingga kebutuhan oksigen meningkat (Hansen, dkk., 2017).

Semakin banyak bahan pencemar yang berada di dalam sungai, maka akan mengakibatkan rendahnya nilai pH yang

membuat kesadahan air yang bersifat asam, sehingga mengindikasikan kualitas air buruk dan tidak layak dipergunakan, sedangkan kadar pH yang sangat tinggi berarti tingkat kandungan basa kuat, di mana zat-zat yang biasa terlarut di dalamnya adalah sisa detergen (Tatangindatu, dkk., 2013).

Kelarutan zat padat dalam air atau disebut dengan TDS adalah terlarutnya zat padat. Tingginya nilai TDS pada air sungai karena banyaknya masyarakat yang menghasilkan limbah domestik dan nondomestik, hal tersebut menyebabkan keadaan air terancam tercemar (Hansen, dkk., 2017).

Konsentrasi amonia dalam air berasal dari air seni dan tinja. Konsentrasi amonia yang tinggi mengindikasikan kebutuhan mikroorganisme terhadap oksigen dalam mengurai bahan organik semakin besar. Konsentrasi amonia yang tinggi menyebabkan konsentrasi BOD juga semakin meningkat (Tatangindatu, dkk., 2013).

Debit adalah volume air per satuan waktu. Dampak yang terjadi apabila debit air tinggi adalah aliran air sungai akan selalu mengalir dengan lancar membawa bahan-bahan organik yang berada di perairan, hal ini menyebabkan sampah tidak akan menumpuk. Bahan-bahan organik yang tidak menumpuk akan membuat mikroorganisme dalam mengurai bahan organik tidak banyak membutuhkan oksigen, sehingga hal ini memengaruhi nilai BOD yang dibutuhkan hanya sedikit (Barid & Yakob, 2007).

Hasil Penelitian dan Pembahasan

Data penelitian ini adalah data sekunder yang diperoleh dari Dinas Lingkungan Hidup Provinsi Kalimantan Timur tahun 2016. Data penelitian terdiri dari data variabel respon, variabel prediktor, dan data koordinat lokasi pengamatan. Data variabel respon (Y) adalah data BOD perairan Sungai Mahakam (golongan air kelas 1) yang dikategorikan menjadi 2 (dikotomis), yaitu $y=1$ jika BOD > 2 mg/l dan $y=0$ jika BOD ≤ 2 mg/l. BOD lebih besar dari 2 mg/l mengindikasikan air Sungai Mahakam tercemar (PP No. 82 Tahun 2001). Data variabel prediktor terdiri dari data suhu (X_1), derajat keasaman atau pH (X_2), Total Dissolved Solids atau TDS (X_3), konsentrasi amonia atau NH_3 (X_4), dan debit air (X_5). Koordinat lokasi pengamatan adalah pasangan letak garis lintang dan bujur dari 27 titik sampel di Daerah Aliran Sungai (DAS) Mahakam Provinsi Kalimantan Timur.

Deskriptif Data Variabel Respon

Deskripsi data variabel respon (data aktual (Y) dan dikotomis (Y_d)) dinyatakan dalam statistik deskriptif yang terdiri dari rata-rata, nilai minimum, nilai maksimum, simpangan baku, dan

rasio status tercemar. Statistik deskriptif disajikan pada Tabel 2.

Tabel 2. Statistik Deskriptif Data Variabel Respon

Respon	Rataan	Min	Maks	Std. Dev	RT (%)
Y	4,88	1,20	13,66	3,99	67,00
Y_d	0,67	0,00	1,00	0,48	67,00

Keterangan: RT = Rasio status tercemar

Berdasarkan Tabel 2, dapat dilihat bahwa persentase lokasi pengamatan terindikasi tercemar sebesar 67% atau 18 dari 27 lokasi pengamatan, di mana nilai rasio ini sama dengan taksiran peluang air Sungai Mahakam tercemar, yaitu sebesar $\hat{\pi} = \bar{Y}_d = 0,67$. Rata-rata BOD adalah 4,88 mg/l, di mana rata-rata ini di atas ambang batas BOD air normal, hal ini mengindikasikan peluang air Sungai Mahakam tercemar cukup besar. BOD tertinggi adalah 13,66 mg/l, yaitu di lokasi pengamatan Karangmumus Hilir dan BOD terendah adalah 1,20 mg/l, yaitu di lokasi pengamatan Belayan Hulu serta Belayan Hilir.

Pendeteksian Multikolinearitas

Pendeteksian multikolinearitas bertujuan untuk mengetahui apakah terdapat hubungan linier antara variabel prediktor dalam suatu model regresi. Nilai VIF dihitung berdasarkan persamaan (15) dan hasil perhitungan dapat dilihat pada Tabel 3.

Tabel 3. Nilai VIF Setiap Variabel Prediktor

Variabel	Nilai VIF
X_1	1,3442
X_2	2,4929
X_3	2,4665
X_4	1,3037
X_5	1,3526

Berdasarkan Tabel 3, dapat disimpulkan bahwa tidak terdapat multikolinearitas antar variabel prediktor. Hal ini ditunjukkan oleh nilai VIF setiap variabel kurang dari 10, sehingga pemodelan regresi logistik biner pada penelitian ini terdiri dari 5 variabel prediktor, yaitu suhu, pH, TDS, konsentrasi amonia, dan debit air.

Model Regresi Logistik Biner Global

Penaksiran parameter model regresi logistik biner global menggunakan metode MLE yang diselesaikan dengan metode iterasi N-R mengacu pada persamaan (10). Hasil penaksiran parameter ditunjukkan pada Tabel 4.

Berdasarkan hasil penaksiran parameter pada Tabel 4 dan berdasarkan model umum persamaan (2), maka model regresi logistik biner global data variabel respon adalah sebagai berikut.

$$\hat{\pi}_i = \frac{\exp\left(-20,7770 + 0,4459X_{i,1} + 1,5648X_{i,2} - 0,0260X_{i,3} + 2,7013X_{i,4} - 0,0019X_{i,5}\right)}{1 + \exp\left(-20,7770 + 0,4459X_{i,1} + 1,5648X_{i,2} - 0,0260X_{i,3} + 2,7013X_{i,4} - 0,0019X_{i,5}\right)}$$

Tabel 4. Penaksiran Parameter Model Regresi Logistik Biner Global

Parameter (β_k)	Nilai Taksiran ($\hat{\beta}_k$)
β_0	-20,7770
β_1	0,4459
β_2	1,5648
β_3	-0,0260
β_4	2,7013
β_5	-0,0019

Ukuran kebaikan model regresi logistik biner global adalah GCV dan R_{MF}^2 . Nilai GCV dan R_{MF}^2 berturut-turut adalah sebesar 0,2166 dan 0,3638.

Pengujian parameter secara simultan bertujuan untuk mengetahui signifikansi parameter regresi terhadap variabel respon secara simultan. Berdasarkan hasil perhitungan pengujian secara simultan, keputusan uji adalah menolak H_0 pada taraf signifikansi 0,10, hal ini ditunjukkan oleh nilai statistik uji $G_{hitung} = 12,5036 > \chi_{(0,10;5)}^2 = 9,2364$ atau $p\text{-value} = 0,0285 < \alpha = 0,10$. Kesimpulan dari uji hipotesis menyatakan bahwa suhu, pH, TDS, konsentrasi amonia, dan debit air secara bersama-sama (simultan) berpengaruh terhadap probabilitas air Sungai Mahakam tercemar.

Pengujian parameter secara parsial bertujuan untuk mengetahui pengaruh setiap variabel prediktor terhadap variabel respon secara individual. Hasil perhitungan pengujian hipotesis parameter secara parsial dapat dilihat pada Tabel 6.

Tabel 6. Pengujian Hipotesis Parameter Regresi Logistik Biner Global Secara Parsial

Variabel	$ W_{hitung} $	P-Value	Keputusan Uji
Konstanta	2,0435	0,0410	H_0 ditolak
X_1	1,6584	0,0972	H_0 ditolak
X_2	2,0875	0,0368	H_0 ditolak
X_3	1,9412	0,0522	H_0 ditolak
X_4	1,7280	0,0840	H_0 ditolak
X_5	1,5871	0,1125	H_0 gagal ditolak

Berdasarkan statistik uji *Wald* yang ditunjukkan pada Tabel 6, disimpulkan bahwa konstanta adalah signifikan. Variabel-variabel prediktor suhu, pH, TDS, dan konsentrasi amonia masing-masing secara individual berpengaruh terhadap probabilitas air Sungai Mahakam

tercemar. Hal ini ditunjukkan dari nilai statistik uji *Wald* keempat variabel tersebut lebih dari 1,64 atau nilai *p-value* masing-masing kurang dari 0,10. Variabel debit air secara individual tidak berpengaruh terhadap probabilitas air Sungai Mahakam tercemar.

Pengujian Heterogenitas Spasial

Pengujian heterogenitas spasial untuk mengetahui apakah data respon dikotomus merupakan data spasial (heterogenitas spasial). Berdasarkan hasil perhitungan $F_{hitung} = 3,7300 > F_{(0,10;5,21)} = 2,1423$ atau $p - value = 0,0142 < \alpha = 0,10$, maka diputuskan menolak H_0 yang berarti data respon dikotomus merupakan data spasial (heterogenitas spasial). Pemodelan regresi logistik biner global diduga tidak sesuai, sehingga pemodelan yang sesuai adalah model *Geographically Weighted Logistic Regression* (GWLRL).

Pemodelan Geographically Weighted Logistic Regression

Langkah pertama dalam penaksiran parameter model GWLRL adalah (1) mencari jarak *Euclidean* antar titik lokasi pengamatan berdasarkan persamaan (17); (2) menentukan *bandwidth* optimum di setiap lokasi pengamatan menggunakan kriteria GCV berdasarkan persamaan (19); (3) menghitung pembobot spasial menggunakan fungsi pembobot *adaptive bisquare* yang diberikan oleh persamaan (18); (4) melakukan penaksiran parameter model GWLRL berdasarkan *bandwidth* yang dicoba, sehingga diperoleh penaksiran parameter pada lokasi pengamatan ke-*i*. Penaksiran parameter model GWLRL setiap lokasi pengamatan diperoleh berdasarkan nilai *bandwidth* yang berbeda (*adaptive*).

Berdasarkan algoritma N-R diperoleh penaksir ML model GWLRL yang disajikan pada Tabel 8.

Tabel 8. Hasil Penaksiran Parameter Model GWLRL

Lokasi	β_0	β_1	β_2	β_3	β_4	β_5
1	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
2	-22,6758	0,4866	1,7587	-0,0299	2,3934	-0,0022
3	-23,6884	0,5052	1,8570	-0,0311	2,2804	-0,0022
4	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
5	-30,8910	0,7077	2,3025	-0,0443	2,1213	-0,0032
6	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
7	-29,5804	0,6978	2,1434	-0,0442	2,1489	-0,0032
8	-30,0339	0,6926	2,2280	-0,0436	2,1177	-0,0031
9	-22,4805	0,4853	1,7288	-0,0298	2,4678	-0,0022
10	-31,9096	0,7194	2,4186	-0,0448	2,0727	-0,0032
11	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
12	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
13	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
14	-23,3711	0,4998	1,8245	-0,0307	2,3231	-0,0022
15	-23,4037	0,4992	1,8327	-0,0307	2,2974	-0,0022
16	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
17	-23,5593	0,5021	1,8475	-0,0309	2,2815	-0,0022
18	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
19	-20,5498	0,4553	1,4997	-0,0271	2,8990	-0,0019
20	-21,7915	0,4922	1,6020	-0,0308	2,7410	-0,0022
21	-23,5309	0,5178	1,8002	-0,0326	2,3982	-0,0024
22	-23,8703	0,5228	1,8368	-0,0328	2,3472	-0,0024
23	-22,6085	0,4869	1,7450	-0,0299	2,4351	-0,0022
24	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019
25	-23,3010	0,4974	1,8223	-0,0306	2,3104	-0,0022
26	-23,3142	0,4983	1,8208	-0,0306	2,3215	-0,0022
27	-20,7770	0,4459	1,5648	-0,0260	2,7013	-0,0019

Pengujian Kesesuaian Model Regresi Logistik Biner Global dan Model Geographically Weighted Logistic Regression

Pengujian kesesuaian model bertujuan untuk mengevaluasi apakah model GWLRL berbeda dari model regresi logistik biner global. Berdasarkan hasil perhitungan, diperoleh bahwa $F_{hitung} = 22,5533 > F_{(0,10;5,135)} = 1,8904$ atau $p - value = 2,2204 \times 10^{-16} < \alpha = 0,10$, maka diputuskan H_0 ditolak dan disimpulkan bahwa model regresi logistik biner global berbeda atau tidak identik dengan model GWLRL.

Ukuran Keباikan Model Regresi Logistik Biner Global dan Model Geographically Weighted Logistic Regression

Perbandingan ukuran kebaikan model regresi logistik biner global dan model GWLRL dapat dilihat pada Tabel 10.

Tabel 10. Ukuran Keباikan Model Regresi Logistik Biner Global dan Model GWLRL

Model	GCV	R_{MF}^2
Regresi Logistik Biner Global	0,2166	0,3638
GWLRL	0,1971	0,5897

Berdasarkan nilai GCV dan R_{MF}^2 yang disajikan pada Tabel 10, model GWLRL lebih baik daripada model regresi logistik biner global karena model GWLRL memiliki nilai GCV yang lebih kecil dan nilai R_{MF}^2 yang lebih besar dibandingkan dengan model regresi logistik biner global.

Pengujian Parameter Model Geographically Weighted Logistic Regression

Pengujian parameter secara simultan bertujuan untuk mengetahui signifikansi parameter regresi terhadap variabel respon secara simultan. Berdasarkan hasil perhitungan, keputusan uji adalah menolak H_0 pada taraf signifikansi 0,10, hal ini ditunjukkan oleh nilai statistik uji

$G_{hitung} = 14,5253 > \chi_{(0,10;7)}^2 = 12,0170$ atau $p - value = 0,0426 < \alpha = 0,10$. Kesimpulan dari uji hipotesis menyatakan bahwa suhu, pH, TDS, konsentrasi amonia, dan debit air secara bersama-sama (simultan) berpengaruh terhadap probabilitas air Sungai Mahakam tercemar.

Pengujian parameter secara parsial bertujuan untuk mengetahui pengaruh variabel prediktor secara individual terhadap variabel respon. Hasil pengujian secara parsial model GWLRL untuk salah satu lokasi pengamatan, yaitu Melak (lokasi pengamatan ke-19) disajikan pada Tabel 12.

Tabel 12. Pengujian Hipotesis Parameter Model GWLR Secara Parsial untuk Lokasi Pengamatan Melak

Variabel	Z _{hitung}	P-Value	Keputusan Uji
Konstanta	1,8815	0,0599	H ₀ ditolak
X ₁	1,5600	0,1138	H ₀ gagal ditolak
X ₂	1,8554	0,0635	H ₀ ditolak
X ₃	1,8602	0,0629	H ₀ ditolak
X ₄	1,6187	0,1055	H ₀ gagal ditolak
X ₅	1,4984	0,1340	H ₀ gagal ditolak

Daerah kritis pengujian hipotesis adalah menolak H₀ pada taraf signifikansi 0,10 jika |Z_{hitung}| > Z_{0,05} = 1,64 atau jika p-value < 0,10. Berdasarkan Tabel 12, variabel prediktor yang berpengaruh secara individual terhadap probabilitas air Sungai Mahakam di lokasi pengamatan Melak (lokasi pengamatan ke-19) tercemar adalah variabel pH dan TDS. Hal ini ditunjukkan dari nilai |Z_{hitung}| yang lebih besar dari 1,64 atau nilai p-value variabel-variabel tersebut lebih kecil dari 0,10.

Berdasarkan hasil penaksiran parameter pada Tabel 8, model GWLR yang menyatakan peluang air Sungai Mahakam di lokasi pengamatan Melak tercemar adalah

$$\hat{\pi}(u_{19}, v_{19}) = \frac{\exp\left(\begin{matrix} -20,5498 + 0,4553X_{19,1} + 1,4997X_{19,2} - \\ 0,0271X_{19,3} + 2,8990X_{19,4} - 0,0019X_{19,5} \end{matrix}\right)}{1 + \exp\left(\begin{matrix} -20,5498 + 0,4553X_{19,1} + 1,4997X_{19,2} - \\ 0,0271X_{19,3} + 2,8990X_{19,4} - 0,0019X_{19,5} \end{matrix}\right)}$$

Berdasarkan hasil pengujian parameter model GWLR secara parsial dan berdasarkan variabel prediktor yang berpengaruh, model GWLR dapat dikelompokkan menjadi 4 kelompok, seperti ditampilkan pada Tabel 13.

Berdasarkan Tabel 13, faktor-faktor (variabel) yang berpengaruh secara lokal terhadap peluang air Sungai Mahakam tercemar adalah suhu, pH, TDS, konsentrasi amonia, dan debit air, sedangkan faktor-faktor yang berpengaruh secara global adalah pH dan TDS.

Model GWLR kelompok pertama adalah model GWLR pada lokasi pengamatan di Kedang Kepala Hulu, Karangmumus Hilir, Kampung Pela, Boh Hulu, Boh Hilir, Bloro, Kantor Gubernur, Anggana, Outlet Danau Semayang, dan Mahakam -Boh. Faktor-faktor yang berpengaruh adalah suhu, pH, TDS, dan konsentrasi amonia. Kelompok kedua merupakan model GWLR pada lokasi pengamatan di Jempang Inlet, Jempang Outlet, Melak, dan Muara Pahu. Faktor-faktor yang berpengaruh adalah pH dan TDS. Model GWLR kelompok ketiga adalah model GWLR pada lokasi pengamatan di Kedang Kepala Hilir, Karangmumus Hulu, Belayan Hilir, Pulau

Kumala, Kalamur, Palaran, Batuq, Muara Muntai, Jembayan, dan Tenggara. Faktor-faktor yang berpengaruh adalah pH, TDS, dan debit air.

Tabel 13. Kelompok Model GWLR Berdasarkan Variabel-Variabel Prediktor yang Berpengaruh

Kel.	Variabel yang Berpengaruh	Daerah Aliran Sungai	Jumlah
1	X ₁ , X ₂ , X ₃ , dan X ₄	Kedang Kepala Hulu Karangmumus Hilir Kampung Pela Boh Hulu Boh Hilir Bloro Kantor Gubernur Anggana Outlet Danau Semayang Mahakam -Boh	10
2	X ₂ dan X ₃	Jempang Inlet Jempang Outlet Melak Muara Pahu	4
3	X ₂ , X ₃ , dan X ₅	Kedang Kepala Hilir Karangmumus Hulu Belayan Hilir Pulau Kumala Kalamur Palaran Batuq Muara Muntai Jembayan Tenggara	10
4	X ₁ , X ₂ , X ₃ , dan X ₅	Kampung Semayang Belayan Hulu Kota Bangun	3

Model GWLR kelompok keempat yang terdiri dari lokasi pengamatan Kampung Semayang, Belayan Hulu, Belayan Hilir, dan Kota Bangun. Faktor-faktor yang berpengaruh adalah suhu, pH, TDS, dan debit air.

Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, maka kesimpulan yang dapat diambil adalah

1. Berdasarkan hasil penaksiran parameter pada Tabel 8, salah satu model GWLR kelompok 1 yang menyatakan peluang air Sungai Mahakam di lokasi pengamatan Kedang Kepala Hulu tercemar, yaitu

$$\hat{\pi}(u_1, v_1) = \frac{\exp\left(\begin{matrix} -20,7770 + 0,4459X_{1,1} + 1,5648X_{1,2} - \\ 0,0260X_{1,3} + 2,7013X_{1,4} - 0,0019X_{1,5} \end{matrix}\right)}{1 + \exp\left(\begin{matrix} -20,7770 + 0,4459X_{1,1} + 1,5648X_{1,2} - \\ 0,0260X_{1,3} + 2,7013X_{1,4} - 0,0019X_{1,5} \end{matrix}\right)}$$

Salah satu model GWLR kelompok 2 yang menyatakan peluang air Sungai Mahakam di lokasi pengamatan Melak tercemar, yaitu

$$\hat{\pi}(u_{19}, v_{19}) = \frac{\exp\left(\begin{matrix} -20,5498 + 0,4553X_{19,1} + 1,4997X_{19,2} - \\ 0,0271X_{19,3} + 2,8990X_{19,4} - 0,0019X_{19,5} \end{matrix}\right)}{1 + \exp\left(\begin{matrix} -20,5498 + 0,4553X_{19,1} + 1,4997X_{19,2} - \\ 0,0271X_{19,3} + 2,8990X_{19,4} - 0,0019X_{19,5} \end{matrix}\right)}$$

Salah satu model GWLR kelompok 3 yang menyatakan peluang air Sungai Mahakam di lokasi pengamatan Muara Muntai tercemar, yaitu

$$\hat{\pi}(u_{22}, v_{22}) = \frac{\exp\left(\begin{matrix} -23,8703 + 0,5228X_{22,1} + 1,8368X_{22,2} - \\ 0,0328X_{22,3} + 2,3472X_{22,4} - 0,0024X_{22,5} \end{matrix}\right)}{1 + \exp\left(\begin{matrix} -23,8703 + 0,5228X_{22,1} + 1,8368X_{22,2} - \\ 0,0328X_{22,3} + 2,3472X_{22,4} - 0,0024X_{22,5} \end{matrix}\right)}$$

Salah satu model GWLR kelompok 4 yang menyatakan peluang air Sungai Mahakam di lokasi pengamatan Kota Bangun tercemar, yaitu

$$\hat{\pi}(u_{23}, v_{23}) = \frac{\exp\left(\begin{matrix} -22,6085 + 0,4869X_{23,1} + 1,7450X_{23,2} - \\ 0,0299X_{23,3} + 2,4351X_{23,4} - 0,0022X_{23,5} \end{matrix}\right)}{1 + \exp\left(\begin{matrix} -22,6085 + 0,4869X_{23,1} + 1,7450X_{23,2} - \\ 0,0299X_{23,3} + 2,4351X_{23,4} - 0,0022X_{23,5} \end{matrix}\right)}$$

2. Faktor-faktor yang berpengaruh bersifat lokal terhadap peluang air Sungai Mahakam tercemar adalah suhu, pH, TDS, konsentrasi amonia, dan debit air, sedangkan faktor-faktor yang berpengaruh secara global adalah pH dan TDS.
3. Interpretasi model GWLR di salah satu lokasi pengamatan, yaitu Melak adalah sebagai berikut. Setiap kenaikan 1 satuan pH dengan asumsi nilai variabel lainnya tetap, akan meningkatkan probabilitas air sungai di lokasi pengamatan Melak tercemar sebesar 4,4806 kali. Setiap kenaikan 1 mg/l TDS dengan asumsi nilai variabel lainnya tetap, akan menurunkan probabilitas air sungai di lokasi pengamatan Melak tercemar sebesar 0,9733 kali.

Daftar Pustaka

Agresti, A. (2002). *Categorical Data Analysis, Second Edition*. New Jersey: John Wiley & Sons, Inc.

Atkinson, P.M., Gemran, S.E., Sear, D.A., & Clark, M.J. (2003). Exploring the Relations between Riverbank Erosion & Geomorphological Controls Using Geographically Weighted Logistic Regression. *Geographically Analysis*. 35(1), 58-82.

Barid, B., & Yakob, M. (2007). Perubahan Kecepatan Aliran Sungai Akibat Perubahan Pelurusan Sungai. *Jurnal Ilmiah Semesta Terbuka*. 10(1), 14-20.

Chasco, C., Garcia, I., & Vicens, J. (2007). Modeling Spatial Variations in Household Disposable Income with Geographically Weighted Regression. *Munich Personal RePEc Archive Paper No. 1682*.

Desriwendi., Hoyyi, A., & Wuryandari, T. (2015). Pemodelan Geographically Weighted Logistic Regression (GWLR) dengan Fungsi Pembobot Fixed Gaussian Kernel dan Adaptive Gaussian Kernel (Studi Kasus: Laju Pertumbuhan Penduduk Provinsi Jawa Tengah). *Jurnal GAUSSIAN*. 4(2), 193-204.

Fathurahman, M., Puhadi., Sutikno., & Ratnasari, V. (2016). Pemodelan Geographically Weighted Logistic Regression pada Indeks Pembangunan Kesehatan Masyarakat di Provinsi Papua. *Prosiding Seminar Nasional MIPA 2016*: 34-42. Jatinangor: Universitas Padjajaran.

Fotheringham, A.S., Brunsdon, C., & Charlton, M.E. (2002). *Geographically Weighted Regression: The Analysis of Spatial Varying Relationships*. England: John Wiley & Sons.

Gujarati, D. (2003). *Ekonometrika Dasar*. Jakarta: Erlangga.

Hansen, Oktaviani, L.W., & Susanti, E.W. (2017). Kualitas Air Sungai Mahakam Terhadap Kesehatan Masyarakat di Kelurahan Loa Duri Kecamatan Loa Janan Kutai Kartanegara Samarinda. *Jurnal Sangkareang Mataram*. 3(4). 17-19.

Kutner, M.H., Nachtsheim, C.J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Model 5th Edition*. New York: McGraw-Hill.

Peraturan Pemerintah Nomor 82 Tahun 2001 Tentang Pengelolaan Kualitas Air dan Pengendalian Pencemaran Air.

Suyitno., Puhadi., Sutikno., & Irhamah. (2016). Parameter Estimation of Geographically Weighted Trivariate Weibull Regression Model. *Applied Mathematical Sciences*. 10(18), 861-878.

Tatangindatu, F., Kalesaran, O., & Rompas, R. (2013). Studi Parameter Fisika Kimia Air pada Areal Budidaya Ikan di Danau Tondano, Desa Paleloan, Kabupaten Minahasa. *Budidaya Perairan*. 1(2). Manado.

Umaly, R.C., & Cuvin, L.A. (1998). *Limnology: Laboratory and Field Guide. Physico-Chemical Factors. Biological Factors*. Philippines: National Bookstore.