

# Pemodelan *Chemical Oxygen Demand* (COD) Sungai di Surabaya Dengan Metode *Mixed Geographically Weighted Regression*

Asih Kurniasih Lumaela, Bambang Widjanarko Otok, dan Sutikno  
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111  
e-mail: sutikno@statistika.its.ac.id

**Abstrak**—*Chemical Oxygen Demand* (COD) merupakan salah satu indikator pencemaran air secara kimia. Surabaya merupakan salah satu kota yang mengalami pencemaran sungai, dimana kondisi sungai di Surabaya memiliki keragaman struktur sehingga memungkinkan terjadinya perbedaan faktor yang berpengaruh pada COD. Untuk mengetahui faktor-faktor yang mempengaruhi COD sungai di Surabaya maka dilakukan pemodelan COD dengan *Mixed Geographically Weighted Regression* (MGWR), metode ini memungkinkan adanya variabel prediktor yang bersifat lokal dan global. Hasil MGWR dengan menggunakan fungsi kernel *Fixed Bisquare* menghasilkan variabel prediktor global yang signifikan adalah Nitrat, sedangkan variabel prediktor lokal yang signifikan adalah kecepatan aliran air dan Nitrit. Namun berdasarkan uji kesesuaian model MGWR, diperoleh kesimpulan bahwa tidak terdapat perbedaan yang signifikan antara model regresi global dan MGWR dengan  $R^2$  sebesar 58,43%. Sehingga pemodelan COD sungai di Surabaya dengan GWR akan memberikan hasil yang lebih baik dibandingkan MGWR, dimana dengan menggunakan GWR diperoleh hasil pemodelan terbaik dengan  $R^2$  sebesar 73,8% pada penggunaan fungsi kernel *Adaptive Bisquare*. Berdasarkan variabel yang signifikan pada model GWR di setiap lokasi diperoleh 9 kelompok, variabel yang signifikan adalah kecepatan aliran air, debit air sungai, Fosfat, Nitrat, Amonia, dan Nitrit.

**Kata kunci**—COD, GWR, MGWR, Fungsi Kernel

## I. PENDAHULUAN

Seiring dengan bertambahnya jumlah penduduk serta semakin berkembangnya industri dan teknologi, seringkali mengakibatkan air mengalami pencemaran. Kota Surabaya merupakan salah satu kota metropolitan dengan tingkat pencemaran sungai yang memprihatinkan. Sekitar 96% air baku Perusahaan Daerah Air Minum (PDAM) Kota Surabaya dipasok dari Kali Surabaya [1]. Kasus pencemaran air di Kali Surabaya mulai terasa pada pertengahan tahun 1976 yang ditandai dengan banyaknya ikan yang mati sehingga PDAM menghentikan produksinya pada saat itu [2]. Pencemaran sungai di Surabaya dapat diketahui melalui jumlah kandungan oksigen yang terlarut dalam air. Salah satu cara yang ditempuh untuk maksud tersebut yaitu dengan uji *Chemical Oxygen Demand* (Wardhana, 2001). *Chemical Oxygen Demand* (COD) merupakan jumlah total oksigen yang dibutuhkan untuk mengoksidasi bahan organik secara kimiawi. Limbah rumah tangga dan industri merupakan sumber utama limbah organik dan merupakan penyebab utama tingginya konsentrasi

COD, selain itu limbah peternakan juga menjadi penyebab tingginya konsentrasi COD [3]. Pada penelitian ini ingin diketahui faktor-faktor yang mempengaruhi tingkat COD pada sungai di Surabaya. Keadaan sungai di Surabaya memiliki perbedaan di setiap lokasi, baik dari segi karakteristik sungai maupun sumber pencemarnya. Kondisi pada setiap lokasi sungai di Surabaya yang berbeda memungkinkan terjadinya perbedaan faktor yang berpengaruh pada nilai COD di setiap lokasi, namun kemungkinan terdapat pula faktor yang memiliki pengaruh di setiap lokasi. Maka salah satu metode yang dapat digunakan untuk melakukan penelitian terkait faktor-faktor yang berpengaruh pada kandungan COD adalah metode *Mixed Geographically Weighted Regression* (MGWR). MGWR merupakan pengembangan dari GWR, metode ini adalah model regresi yang beberapa koefisien dari variabel indikatornya bersifat konstan, sedangkan yang lainnya bervariasi secara spasial [4]. Penelitian dengan menggunakan metode MGWR juga telah dilakukan oleh [5] untuk mengetahui faktor-faktor yang mempengaruhi persentase rumah tangga miskin di Kabupaten Mojokerto tahun 2008. Dengan metode ini, mampu memodelkan rumah tangga miskin lebih baik daripada metode GWR.

## II. TINJAUAN PUTSAKA

### A. Regresi Linear

Analisis regresi merupakan metode statistika yang digunakan untuk menentukan hubungan antara variabel respon ( $y$ ) dan variabel prediktor ( $x$ ), dalam regresi berganda dapat diprediksi variabel respon berdasarkan asumsi hubungan linear dengan beberapa variabel prediktor  $x_1, x_2, \dots, x_k$  [6]. Model regresi linear berganda untuk  $k$  variabel prediktor dapat ditulis dalam bentuk persamaan (1) berikut.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (1)$$

### B. *Geographically Weighted Regression*

*Geographically Weighted Regression* (GWR) adalah teknik regresi lokal yang memungkinkan parameter model bervariasi di setiap lokasi [7]. Parameter model regresi di setiap lokasi akan menghasilkan nilai yang berbeda-beda. Terdapat observasi  $\{x_{ip}\}$  untuk  $i=1,2,\dots,n$  pengamatan dan  $p=1,2,\dots,k$  variabel prediktor, serta variabel respon  $\{y_i\}$  dan koordinat lokasi  $\{(u_i, v_i)\}$  untuk setiap pengamatan. Model GWR diberikan oleh persamaan (2) berikut.

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^k \beta_p(u_i, v_i)x_{ip} + \varepsilon_i \quad (2)$$

$\varepsilon_i$  adalah error pada pengamatan ke- $i$  dengan asumsi dasar identik, independen, dan berdistribusi normal dengan rata-rata nol dan varians  $\sigma^2$  [8]. Estimasi parameter model GWR menggunakan *Weighted Least Square* (WLS) dapat ditulis sebagai berikut.

$$\hat{\beta}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y}$$

Dimana  $\mathbf{W}(u_i, v_i)$  merupakan matriks diagonal dengan elemennya adalah  $w_j(u_i, v_i)$  yang merupakan bobot untuk data pada titik  $j$  dalam pengujian model di sekitar titik  $(u_i, v_i)$  dengan  $j = 1, 2, \dots, n$ . Besarnya pembobotan untuk model GWR di setiap lokasi dapat ditentukan dengan menggunakan fungsi kernel, dimana terdapat fungsi kernel *fixed* dan *adaptive*. Fungsi kernel *fixed* memiliki nilai *bandwidth* yang sama untuk semua lokasi pengamatan, fungsi pembobot dari fungsi kernel *fixed* dapat ditulis sebagai berikut.

1. Jarak Gaussian [7]

$$w_j(u_i, v_i) = \sqrt{\exp(-(d_{ij}/h)^2)}$$

2. Bisquare [9]

$$w_j(u_i, v_i) = \begin{cases} \left[1 - (d_{ij}/h)^2\right]^2 & , \text{ untuk } d_{ij} < h \\ 0 & , \text{ untuk lainnya} \end{cases}$$

3. Tricube [9]

$$w_j(u_i, v_i) = \begin{cases} \left[1 - (d_{ij}/h)^3\right]^3 & , \text{ untuk } d_{ij} < h \\ 0 & , \text{ untuk lainnya} \end{cases}$$

Sedangkan fungsi kernel *adaptive* memiliki *bandwidth* yang berbeda-beda di setiap lokasi pengamatan, dengan fungsi pembobot sebagai berikut.

1. Jarak Gaussian

$$w_j(u_i, v_i) = \sqrt{\exp(-(d_{ij}/h_i)^2)}$$

2. Bisquare

$$w_j(u_i, v_i) = \begin{cases} \left[1 - (d_{ij}/h_i)^2\right]^2 & , \text{ untuk } d_{ij} < h_i \\ 0 & , \text{ untuk lainnya} \end{cases}$$

3. Tricube

$$w_j(u_i, v_i) = \begin{cases} \left[1 - (d_{ij}/h_i)^3\right]^3 & , \text{ untuk } d_{ij} < h_i \\ 0 & , \text{ untuk lainnya} \end{cases}$$

$h_i$  pada fungsi kernel *adaptive* merupakan *bandwidth* pada lokasi pengamatan ke- $i$ .

Pada ketiga fungsi kernel di atas,  $d_{ij}$  adalah jarak antara titik di lokasi  $i$  dan lokasi  $j$  yang dapat dicari dengan jarak *euclidean*  $d_{ij} = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2}$ . Sementara  $h$  adalah parameter non negatif yang dikenal sebagai *bandwidth* atau parameter penghalus. Nilai *bandwidth* ditentukan dengan menggunakan prosedur *Cross Validation* (CV) dengan bentuk matematis seperti persamaan (3) [10].

$$CV(h) = \sum_{i=1}^n [y_i - \hat{y}_{zi}(h)]^2 \quad (3)$$

$\hat{y}_{zi}(h)$  adalah nilai taksiran untuk  $y_i$ , dengan menghilangkan observasi pada titik  $i$  pada proses pengujian parameter.

### C. Mixed Geographically Weighted Regression

Metode MGWR digunakan pada saat terdapat variabel prediktor tertentu yang berpengaruh pada variabel respon secara global serta terdapat pula variabel prediktor yang berpengaruh secara lokal. Model MGWR untuk  $k$  variabel prediktor dengan  $q$  variabel prediktor diantaranya yang bersifat lokal serta dimisalkan intersep model bersifat lokal dapat ditulis sebagai berikut [5].

$$y_i = \beta_0(u_i, v_i) + \sum_{p=1}^q \beta_p(u_i, v_i)x_{ip} + \sum_{p=q+1}^k \beta_p x_{ip} + \varepsilon_i \quad i=1,2,\dots,n$$

Estimasi parameter untuk model MGWR dapat diperoleh dengan menggunakan metode *Weighted Least Square*, untuk estimasi parameter yang bersifat global adalah :

$$\hat{\beta}_g = [\mathbf{X}_g^T (\mathbf{I} - \mathbf{S}_l)^T (\mathbf{I} - \mathbf{S}_l) \mathbf{X}_g]^{-1} \mathbf{X}_g^T (\mathbf{I} - \mathbf{S}_l)^T (\mathbf{I} - \mathbf{S}_l) \mathbf{y}$$

Sedangkan untuk estimasi parameter yang bersifat lokal adalah :

$$\hat{\beta}_l(u_i, v_i) = [\mathbf{X}_l^T \mathbf{W}(u_i, v_i) \mathbf{X}_l]^{-1} \mathbf{X}_l^T \mathbf{W}(u_i, v_i) (\mathbf{Y} - \mathbf{X}_g \hat{\beta}_g) \quad \text{untuk}$$

$i=1,2,\dots,n$ . dengan  $\mathbf{S}_g = \mathbf{X}_g (\mathbf{X}_g^T \mathbf{X}_g)^{-1} \mathbf{X}_g^T$  [11].

Hipotesis pengujian kesesuaian model regresi global dan MGWR untuk mengetahui faktor geografis yang signifikan adalah :

$H_0 : \beta_p(u_i, v_i) = \beta_p$  untuk  $p=0,1,2, \dots, q$  dan  $i=1,2, \dots, n$

$H_1 : \text{minimal terdapat satu } \beta_p(u_i, v_i) \neq \beta_p$

Statistik uji :

$$F(1) = \frac{(\mathbf{y}^T [(\mathbf{I} - \mathbf{H}) - (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})] \mathbf{y}) / v_1}{(\mathbf{y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{y}) / u_1} \quad (4)$$

dengan

$$\mathbf{S} = \mathbf{S}_l + (\mathbf{I} - \mathbf{S}_l) \mathbf{X}_g [\mathbf{X}_g^T (\mathbf{I} - \mathbf{S}_l)^T (\mathbf{I} - \mathbf{S}_l) \mathbf{X}_g]^{-1} \mathbf{X}_g^T (\mathbf{I} - \mathbf{S}_l)^T (\mathbf{I} - \mathbf{S}_l)$$

$H_0$  ditolak jika  $F(1) \geq F_{\alpha; df_1, df_2}$ , dengan  $df_1 = v_1^2 / v_2$  dan

$df_2 = u_1^2 / u_2$ . Dimana  $v_i = \text{tr}([\mathbf{I} - \mathbf{H}) - (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})]^i)$  dan

$u_i = \text{tr}([\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})]^i)$  untuk  $i=1,2,\dots,n$ .

Uji serentak pada variabel prediktor yang bersifat global dilakukan dengan hipotesis :

$H_0 : \beta_{q+1} = \beta_{q+2} = \dots = \beta_k = 0$

$H_1 : \text{minimal ada satu } \beta_p \neq 0$ , untuk  $p = q+1, q+2, \dots, k$

Statistik uji :

$$F(2) = \frac{(\mathbf{y}^T [(\mathbf{I} - \mathbf{S}_l)^T (\mathbf{I} - \mathbf{S}_l) - (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S})] \mathbf{y}) / r_1}{(\mathbf{y}^T (\mathbf{I} - \mathbf{S})^T (\mathbf{I} - \mathbf{S}) \mathbf{y}) / u_1}$$

Tolak  $H_0$  jika  $F(2) \geq F_{\alpha;df_1,df_2}$ , dimana  $df_1 = r_1^2/r_2$ ,

$$df_2 = \frac{u_1^2}{u_2}, \& r_i = tr\left(\left[(\mathbf{I}-\mathbf{S}_i)^T(\mathbf{I}-\mathbf{S}_i)-(\mathbf{I}-\mathbf{S})^T(\mathbf{I}-\mathbf{S})\right]^i\right), i=1,2.$$

Hipotesis untuk uji serentak pada variabel prediktor yang bersifat lokal  $x_p(1 \leq p \leq q)$  adalah:

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

$$H_1 : \text{minimal terdapat satu } \beta_p(u_i, v_i) \neq 0, p = 1, 2, \dots, q$$

Statistik uji :

$$F(3) = \frac{\left(\mathbf{y}^T \left[ (\mathbf{I}-\mathbf{S}_g)^T (\mathbf{I}-\mathbf{S}_g) - (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S}) \right] \mathbf{y}\right) / t_1}{\left(\mathbf{y}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S}) \mathbf{y}\right) / u_1}$$

Keputusan tolak  $H_0$  jika  $F(3) \geq F_{\alpha;df_1,df_2}$ . dimana

$$df_1 = \frac{t_1^2}{t_2}, df_2 = \frac{u_1^2}{u_2}, t_i = tr\left(\left[(\mathbf{I}-\mathbf{S}_g)^T (\mathbf{I}-\mathbf{S}_g) - (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\right]^i\right)$$

Uji parsial untuk mengetahui parameter variabel prediktor global  $x_p(q+1 \leq p \leq k)$  yang berpengaruh secara signifikan terhadap variabel respon dilakukan dengan hipotesis :

$$H_0 : \beta_p = 0 \text{ (variabel global } x_p \text{ tidak signifikan)}$$

$$H_1 : \beta_p \neq 0 \text{ (variabel global } x_p \text{ signifikan)}$$

$$\text{Statistik uji : } T_g = \frac{\hat{\beta}_p}{\hat{\sigma} \sqrt{g_{pp}}}$$

Dimana  $g_{pp}$  adalah elemen diagonal ke- $p$  dengan  $q+1 \leq p \leq k$  dari matriks  $\mathbf{G}\mathbf{G}^T$ , dimana

$$\mathbf{G} = \left[ \mathbf{X}_g^T (\mathbf{I}-\mathbf{S}_i)^T (\mathbf{I}-\mathbf{S}_i) \mathbf{X}_g \right]^{-1} \mathbf{X}_g^T (\mathbf{I}-\mathbf{S}_i)^T (\mathbf{I}-\mathbf{S}_i)$$
 dan

$$\hat{\sigma}^2 = \frac{\mathbf{y}^T (\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S}) \mathbf{y}}{tr\left((\mathbf{I}-\mathbf{S})^T (\mathbf{I}-\mathbf{S})\right)}$$

Tolak  $H_0$  jika  $|T_{g\_hitung}| > t_{\alpha/2;df_i}$  dengan  $df = u_1^2/u_2$ .

Sedangkan uji parsial untuk mengetahui pengaruh signifikan parameter variabel prediktor yang bersifat lokal  $x_p(1 \leq p \leq q)$  dilakukan dengan hipotesis sebagai berikut.

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

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

$$\text{Statistik uji : } T_i = \frac{\hat{\beta}_p(u_i, v_i)}{\hat{\sigma} \sqrt{m_{pp}}}$$

Dimana  $m_{pp}$  adalah elemen diagonal ke- $p$  dengan  $1 \leq p \leq q$  dari matriks  $\mathbf{M}\mathbf{M}^T$ , dengan

$$\mathbf{M} = \left[ \mathbf{X}_i^T \mathbf{W}(u_i, v_i) \mathbf{X}_i \right]^{-1} \mathbf{X}_i^T \mathbf{W}(u_i, v_i) (\mathbf{I} - \mathbf{X}_g \mathbf{G})$$
. Tolak  $H_0$

jika  $|T_{i\_hitung}| > t_{\alpha/2;df}$  dengan  $df = u_1^2/u_2$ .

#### D. Chemical Oxygen Demand

Chemical Oxygen Demand atau kebutuhan oksigen kimia adalah jumlah oksigen yang diperlukan agar bahan buangan yang ada di dalam air dapat teroksidasi melalui reaksi kimia. Sumber COD berasal dari kegiatan industri kertas, penyamakan kulit, gula, pemotongan daging, pengalengan ikan, pembekuan udang, roti, susu, keju, dan mentega, limbah domestik dan lain-lain. Keberadaan COD di lingkungan akan

memberikan dampak pada manusia dan lingkungan, diantaranya adalah banyaknya biota air yang mati karena konsentrasi oksigen terlarut dalam air terlalu sedikit dan semakin sulitnya mendapatkan air sungai yang memenuhi kriteria sebagai bahan baku air minum [2].

### III. METODE PENELITIAN

#### A. Data Penelitian

Data pada penelitian ini adalah data sekunder yang diperoleh dari Balai Besar Teknik Kesehatan Lingkungan dan Pengendalian Penyakit (BBTKLPP) Surabaya dan Badan Lingkungan Hidup Kota Surabaya. Sedangkan data karakteristik lokasi sungai diperoleh dari hasil penelitian [12]. Unit penelitian yang diteliti adalah sungai di Surabaya yang meliputi kali, saluran air, dan boezem, dengan jumlah sampel ( $n$ ) sebanyak 25 titik lokasi pengamatan sungai di Surabaya. Variabel respon ( $Y$ ) penelitian ini adalah COD, sedangkan variabel prediktor yang digunakan adalah lebar sungai ( $X_1$ ), kedalaman sungai ( $X_2$ ), kecepatan aliran air ( $X_3$ ), debit air ( $X_4$ ), Fosfat ( $X_5$ ), Nitrat ( $X_6$ ), Amonia ( $X_7$ ), Nitrit ( $X_8$ ), suhu air ( $X_9$ ), dan kandungan detergen ( $X_{10}$ ). Variabel prediktor Fosfat, Nitrat, Amonia, Nitrit, dan detergen merupakan pendekatan dari sumber COD yaitu limbah rumah tangga, limbah industri, dan limbah peternakan. Detergen dan Nitrat yang bersumber dari limbah rumah tangga, sedangkan untuk limbah peternakan dilakukan pendekatan dengan sumber pencemar Fosfat. Sementara Nitrit dan Amonia merupakan pendekatan untuk limbah industri, dimana amonia bersumber dari industri penyamakan kulit, pengalengan ikan, dan lain-lain.

#### B. Langkah Analisis

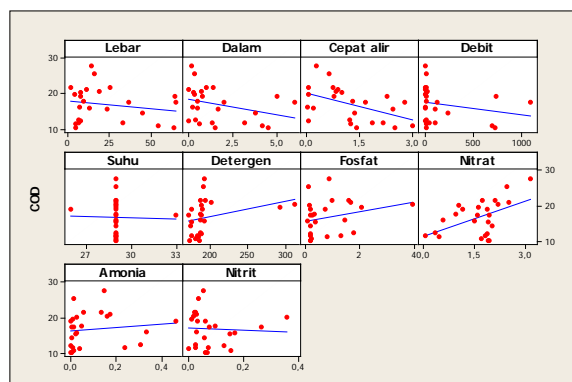
Langkah-langkah yang dilakukan setelah menentukan variabel prediktor yang diperkirakan mempengaruhi COD dalam air sungai adalah sebagai berikut.

1. Mengidentifikasi pola hubungan antara variabel prediktor dan variabel respon, serta menguji multikolinearitas.
2. Melakukan pemodelan regresi linear global yang meliputi estimasi parameter, pengujian kesesuaian model, pengujian parsial parameter regresi, dan pengujian asumsi residual IIDN (identik, independen, dan berdistribusi normal).
3. Melakukan pemodelan GWR yang diawali dengan menghitung jarak *euclidean* antara lokasi  $i$  dengan koordinat  $(u_i, v_i)$  dan  $j$  dengan koordinat  $(u_j, v_j)$ . Kemudian menentukan *bandwidth* optimum dengan metode *Cross Validation* (CV). Setelah dihitung besarnya pembobot model GWR dilakukan estimasi parameter, pengujian kesesuaian model dan uji parsial.
4. Melakukan pemodelan MGWR yang diawali dengan menentukan variabel prediktor global dan variabel prediktor lokal, kemudian dilakukan estimasi parameter, uji kesesuaian model serta uji serentak dan uji parsial parameter dari variabel prediktor lokal dan parameter variabel prediktor global.
5. Menentukan model terbaik dengan menggunakan AIC, dimana nilai dari AIC untuk model regresi linear, GWR,

Tabel 1.

Nilai Rataan, Standar Deviasi, Minimum, dan Maksimum Variabel Penelitian

Variabel	Rataan	Std.dev.	Min.	Max.
COD (mg/l)	16,897	4,789	10,264	27,619
Lebar (m)	22,240	20,970	2	65,200
Kedalaman (m)	1,634	1,750	0,050	6
Cepat alir (m/dtk)	1,227	0,915	0,030	3
Debit (m <sup>3</sup> /dtk)	155,3	304,000	0,023	1087,500
Fosfat (mg/l)	0,781	0,910	0,064	3,937
Nitrat (mg/l)	1,613	0,729	0,047	3,149
Amonia (mg/l)	0,091	0,126	0,003	0,465
Nitrit (mg/l)	0,081	0,084	0,001	0,356
Suhu (°C)	29,040	1,020	26	33
Detergen (mg/l)	194,280	33,500	172	312



Gambar 1. Diagram Pencar antara Nilai COD dan Variabel Prediktor

6. dan MGWR dibandingkan. Model terbaik adalah model yang memiliki nilai AIC terkecil.

IV. ANALISIS DAN PEMBAHASAN

A. Deskripsi Variabel COD dan Faktor-Faktor yang Mempengaruhi

Tabel 1 menunjukkan bahwa rata-rata kadar COD sungai di Surabaya adalah 16,897 mg/l, nilai ini melebihi kriteria kadar COD air sungai untuk digunakan sebagai air minum yaitu sebesar 10 mg/l. Kadar COD tertinggi adalah 27,619 mg/l yang terletak di Kali Wonorejo, sedangkan kadar COD terendah sebesar 10,264 mg/l terletak pada Kali Surabaya di Kedurus. Lebar sungai dan debit air sungai yang diamati memiliki variasi yang besar antar sungai. Lebar sungai memiliki variasi 20,97 meter, sedangkan debit sungai memiliki variasi antar lokasi sebesar 304 m<sup>3</sup>/dtk. Ditinjau dari nilai kandungan Fosfat, Nitrat, dan Amonia tertinggi yang secara berurutan terdapat pada Kali Greges, Kali Wonorejo, dan Kali Kepiting, diketahui bahwa ketiga sungai tersebut terdapat pada lokasi yang padat pemukiman penduduk dan telah mendekati hilir sungai sehingga kemungkinan jumlah senyawa kimia pada ketiga sungai tersebut telah terakumulasi.

B. Identifikasi Pola Hubungan COD dan Faktor-Faktor yang Mempengaruhinya

Sebelum memodelkan nilai COD terlebih dahulu dilakukan identifikasi pola hubungan dari sepuluh variabel prediktor yang ada terhadap nilai COD. Gambar 1 berikut ini menunjukkan pola hubungan COD dan setiap variabel prediktor dengan diagram pencar.

Berdasarkan Gambar 1 dapat diketahui bahwa variabel

Tabel 2.

Perbandingan Fungsi Kernel GWR

	Gaussian		Bisquare		Tricube	
	Fixed	Adapt.	Fixed	Adapt.*	Fixed	Adapt.
AIC	136,87	130,39	136,46	128,13	136,66	129,16
R <sup>2</sup>	0,561	0,711	0,585	0,738	0,578	0,722
SSE	241,64	159,36	228,55	144,20	232,53	152,98

Keterangan :\*) Model GWR terbaik

prediktor yang memiliki hubungan positif dengan kandungan COD dalam air sungai adalah Fosfat dan Nitrat, artinya jika terdapat peningkatan kandungan Fosfat dan Nitrat dalam air sungai maka akan mengakibatkan peningkatan COD di sungai tersebut. Selain itu untuk variabel prediktor suhu dan detergen terdapat kecenderungan untuk tidak berpengaruh pada nilai COD, hal ini terlihat dari hasil plot yang tidak menyebar namun relatif tersebar di satu titik saja. Oleh karena itu variabel suhu dan detergen tidak akan digunakan dalam analisis selanjutnya.

C. Regresi Linear

Nilai dari variabel prediktor distandardisasi terlebih dahulu karena satuan dari variabel prediktor tidak sama. Variabel prediktor yang sudah distandardkan (variabel Z) merupakan variabel prediktor yang dikurangi dengan nilai mean dan dibagi dengan standar deviasinya. Dengan menggunakan metode regresi *stepwise* yang menghasilkan dua variabel prediktor yang signifikan yaitu kecepatan aliran air (x<sub>3</sub>) dan kandungan Nitrat dalam air sungai (x<sub>6</sub>). Sehingga diperoleh persamaan untuk regresi global adalah sebagai berikut.

$$\hat{y} = 16,9 - 2,23Z_3 + 2,34Z_6 \tag{5}$$

Berdasarkan persamaan (5) dapat diketahui bahwa variabel prediktor yang berpengaruh pada nilai COD sungai di Surabaya adalah cepat alir dan kandungan Nitrat.

Pengujian asumsi residual dari model regresi global menunjukkan bahwa residual telah berdistribusi normal, independen, dan tidak identik. Asumsi residual identik tidak terpenuhi, sehingga diduga bahwa nilai COD sungai di Surabaya menyebar secara spasial titik dan dilanjutkan analisis dengan metode GWR.

D. Geographically Weighted Regression

Terdapat tiga fungsi kernel yang akan dibandingkan untuk model GWR yaitu Gaussian, Bisquare, dan Tricube. Berdasarkan hasil perbandingan ketiga fungsi kernel tersebut yang terdapat pada Tabel 2, dapat diketahui bahwa fungsi kernel *Adaptive* Bisquare memberikan nilai AIC terkecil, R<sup>2</sup> terbesar, dan SSE terkecil dibandingkan dengan fungsi kernel yang lain. Sehingga disimpulkan bahwa *Adaptive* Bisquare memberikan model GWR terbaik dan akan digunakan untuk menentukan pembobotan pada model GWR.

Pengujian kesesuaian model GWR dengan statistik uji F dilakukan untuk mengetahui apakah terdapat perbedaan yang signifikan antara model regresi linear dan GWR. Diperoleh hasil uji F untuk kesesuaian model GWR dengan *P-value* 0,1923, sehingga dengan menggunakan tingkat signifikansi (α) 20% dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara model regresi linear dan GWR dalam

Tabel 3.

Pengelompokan Sungai Berdasarkan Variabel yang Signifikan pada  $\alpha=20\%$  dengan GWR

Nama Sungai	Variabel
Kali Jeblokan di Jl.Kedung Cowek, Kali Wonorejo, Kali Greges	$x_6$
Kali Surabaya di Karangpilang, Kali Surabaya di DAM Gunungsari	$x_3, x_7$
Kali Surabaya di DAM Jagir	$x_3, x_8$
Kali Wonokromo, Kali Kebon Agung	$x_6, x_7$
Kali Surabaya di Kedurus	$x_7, x_8$
Kali Banyu Urip	$x_5, x_8$
Kali Dami, Kali Kepiting, Boezem Kali Dami, Boezem Wonorejo	$x_4, x_6$
Kali Surabaya di Wonokromo, Kali Mas di Ngagel, Kali Bokor, Saluran Darmo	$x_6, x_7, x_8$
Kali Mas di Keputran Selatan, Kali Mas di Kebon Rojo, Kali Jeblokan di Jl.Petojo, Kali Pergirikan di Jl.Undaan, Kali Pergirikan di Jl.Pergirikan, Saluran Dinoyo, Saluran Kenari	$x_6, x_8$

memodelkan COD sungai di Surabaya.

Langkah selanjutnya adalah menguji parsial setiap parameter pada masing-masing lokasi sungai di Surabaya untuk mengetahui variabel apa saja yang berpengaruh signifikan dalam model GWR. Hasil pengelompokan sungai di Surabaya berdasarkan variabel prediktor yang signifikan pada tingkat signifikansi ( $\alpha$ ) sebesar 20% ditunjukkan pada Tabel 3 berikut ini.

Berdasarkan hasil pengelompokan variabel prediktor yang signifikan pada  $\alpha$  sebesar 20%, kemudian dapat dibentuk model GWR untuk setiap lokasi. Salah satu model yang terbentuk untuk Kali Surabaya di Wonokromo ditunjukkan oleh persamaan berikut ini.

$$\hat{y}_2 = 8 + 3,64x_6 + 12,08x_7 + 15x_8 \quad (6)$$

Model GWR pada persamaan (6) tersebut menjelaskan bahwa COD Kali Surabaya di Wonokromo akan naik sebesar 3,64 mg/l jika kandungan Nitrat dalam air meningkat sebesar 1 mg/l dengan syarat variabel prediktor yang lain bersifat konstan. Meningkatnya kandungan Nitrat yang berasal dari limbah rumah tangga menyebabkan air menjadi tercemar dimana kadar bahan buangan oraganik maupun non-oraganik dalam air meningkat, sehingga jumlah oksigen yang dibutuhkan untuk mengoksidasi bahan buangan dalam air sungai secara kimia atau yang biasa disebut COD ikut meningkat.

#### E. Mixed Geographically Weighted Regression

Hasil pemodelan GWR untuk setiap lokasi dengan tingkat signifikansi sebesar 20% pada Tabel 3 menunjukkan bahwa variabel Nitrat ( $x_6$ ) signifikan hampir di semua lokasi, selain itu meninjau dari sumbernya Nitrat berasal dari limbah rumah tangga dimana hampir di sepanjang sungai di Surabaya terdapat banyak pemukiman penduduk yang menjadi sumber dari limbah rumah tangga. Mempertimbangkan dua kondisi tersebut maka diduga variabel prediktor Nitrat memiliki pengaruh global di semua lokasi pengamatan sungai di Surabaya, sehingga analisis dilanjutkan dengan menggunakan metode MGWR.

Fungsi kernel *Adaptive* Bisquare pada pemodelan GWR

Tabel 4.

Perbandingan Fungsi Kernel MGWR

	Gaussian		Bisquare		Tricube	
	Fixed	Adapt.	Fixed*	Adapt.	Fixed	Adapt.
AIC	166,13	186,71	150,00	207,25	149,87	219,27
R <sup>2</sup>	0,095	-	0,5843	-	0,5775	-
SSE	498,13	876,12	228,82	755,77	232,53	4615,73

Ket : \*) Model MGWR terbaik ; -) R<sup>2</sup> bernilai negatif

Tabel 5.

Nilai Statistik Uji F Model MGWR

	F hitung	df1	df2	P-value
F1	0,7948992	7,005406	22,83471	0,599
F2	7,0066301	8,904179	22,83471	0,000
F3	49,408381	20,018884	22,83471	0,000

memberikan hasil AIC yang terbaik dibandingkan fungsi kernel lainnya, namun pada pemodelan MGWR fungsi kernel *Adaptive* Bisquare memberikan hasil AIC yang besar. Selain itu pada Tabel 4 juga dapat diketahui bahwa tidak terdapat hasil R<sup>2</sup> untuk semua fungsi kernel *adaptive*. Hal ini dikarenakan nilai SSE yang diperoleh lebih besar dibandingkan SST, akibatnya nilai R<sup>2</sup> menjadi negatif sehingga nilai R<sup>2</sup> tersebut tidak valid dan tidak dituliskan. Besarnya nilai SSE dibandingkan SST terjadi dikarenakan terdapat matriks singular pada operasi perkalian matriks  $\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}_l$  di beberapa lokasi pengamatan.

Untuk itu sebelum analisis MGWR dilanjutkan maka terlebih dahulu dipilih kembali fungsi kernel yang akan digunakan untuk membentuk matriks pembobot. Berdasarkan Tabel 4 diketahui bahwa fungsi kernel yang akan digunakan adalah *Fixed* Bisquare. Meskipun nilai AIC model MGWR dengan fungsi kernel *Fixed* Tricube lebih kecil dari *Fixed* Bisquare, namun selisih kedua nilai AIC tersebut tidak berbeda jauh sehingga jika dilihat dari nilai SSE fungsi kernel *Fixed* Bisquare lebih kecil dibandingkan fungsi kernel lainnya.

Nilai *bandwidth* yang dihasilkan pada fungsi kernel *Fixed* Bisquare akan bernilai sama di setiap lokasi pengamatan sungai di Surabaya, dimana berdasarkan hasil pengolahan data dengan menggunakan program R diperoleh nilai *bandwidth* ( $h$ ) yang digunakan adalah 0,7945013.

Tabel 5 menunjukkan hasil statistik uji F, dimana F1 menunjukkan statistik uji untuk pengujian kesesuaian model MGWR, kemudian F2 menunjukkan hasil pengujian serentak variabel prediktor yang bersifat global, dan F3 untuk uji serentak variabel prediktor yang bersifat lokal. Sehingga berdasarkan *P-value* F1 disimpulkan bahwa tidak terdapat perbedaan yang signifikan antara model MGWR dan regresi global, sedangkan untuk hasil F2 dan F3 menunjukkan bahwa terdapat minimal satu variabel prediktor baik global maupun lokal yang signifikan dalam model MGWR.

Pada pemodelan COD sungai di Surabaya dengan MGWR hanya terdapat satu variabel prediktor yang diperkirakan bersifat global di seluruh titik pengamatan yaitu Nitrat. Uji parsial variabel prediktor global dilakukan dengan statistik uji  $t$ , dengan menggunakan tingkat signifikansi ( $\alpha$ ) sebesar 5% dan  $df=22,835$  diperoleh kesimpulan bahwa Nitrat

Tabel 6.

Pengelompokan Sungai Berdasarkan Variabel Lokal yang Signifikan pada  $\alpha=20\%$  dengan MGWR

Nama Sungai	Variabel
Kali Mas di Kebon Rojo, Kali Jeblokan di Jl.Petojo, Kali Jeblokan di Jl.Kedung Cowek, Kali Pergirikan di Jl.Pergirikan, Kali Dami, Kali Bokor, Kali Wonorejo, Kali Kepiting, Kali Wonokromo, Boezem Kali Dami, Boezem Wonorejo, Kali Kebon Agung, Kali Surabaya di Karangpilang, Kali Surabaya di DAM Gunungsari, Kali Banyu Urip, Saluran Darmo, Kali Mas di Keputran Selatan, Kali Pergirikan di Jl.Undaan, Kali Greges, Saluran Dinoyo, Saluran Kenari, Kali Surabaya di Kedurus, Kali Surabaya di Wonokromo, Kali Mas di Ngagel	$x_3$
Kali Surabaya di DAM Jagir	$x_3, x_8$

Tabel 7.

Perbandingan Model Regresi Global, GWR, dan MGWR

Metode		AIC	SSE	R <sup>2</sup>
Regresi Global		137,18	246,69	0,552
GWR	Gaussian <i>Fixed</i>	136,87	241,64	0,561
	Gaussian <i>Adaptive</i>	130,39	159,36	0,7105
	Bisquare <i>Fixed</i>	136,46	228,55	0,5847
	Bisquare <i>Adaptive*</i>	128,13	144,20	0,738
	Tricube <i>Fixed</i>	136,66	232,53	0,5775
	Tricube <i>Adaptive</i>	129,16	152,98	0,722
MGWR	Gaussian <i>Fixed</i>	166,13	498,13	0,095
	Gaussian <i>Adaptive</i>	186,71	876,12	-
	Bisquare <i>Fixed</i>	150,00	228,82	0,5843
	Bisquare <i>Adaptive</i>	207,25	755,77	-
	Tricube <i>Fixed</i>	149,87	232,53	0,5775
	Tricube <i>Adaptive</i>	219,27	4615,73	-

Ket : \*) Model terbaik ; -) R<sup>2</sup> bernilai negatif

berpengaruh secara signifikan pada COD secara global di semua titik pengamatan sungai. Estimasi parameter untuk Nitrat pada pemodelan MGWR ini adalah 2,421.

Berdasarkan hasil uji parsial untuk variabel prediktor lokal, diperoleh dua kelompok yang terbentuk pada tingkat signifikan 20% seperti pada Tabel 6. Jika ditinjau variabel prediktor lokal yang signifikan adalah kecepatan aliran air ( $x_3$ ) dan Nitrit ( $x_8$ ), dimana variabel kecepatan aliran air berpengaruh pada semua lokasi.

Salah satu model MGWR yang terbentuk untuk Kali Surabaya di DAM Jagir ditunjukkan pada persamaan (7) berikut.

$$\hat{y}_{25} = 4,95 - 3,593x_3 + 3,217x_6 + 26,235x_8 \quad (7)$$

Melalui persamaan (4) dapat diketahui bahwa jika kecepatan aliran air ( $x_3$ ) Kali Surabaya di DAM Jagir naik 1 m/dtk maka COD akan berkurang 3,593 mg/l, dengan syarat variabel prediktor lainnya konstan. Kecepatan aliran air berpengaruh pada penurunan COD dikarenakan ketika kecepatan aliran air meningkat maka akan mengakibatkan oksigen yang ada pada perairan semakin banyak, meningkatnya oksigen pada air akan membuat COD semakin berkurang.

#### A. Pemilihan Model Terbaik

Ringkasan nilai AIC, R<sup>2</sup>, dan SSE dari metode regresi global, GWR, dan MGWR ditunjukkan pada Tabel 7.

Berdasarkan hasil ringkasan pemodelan COD sungai di Surabaya pada Tabel 6 dengan menggunakan tiga metode yaitu regresi global, GWR, dan MGWR, baik dengan fungsi kernel *fixed* maupun *adaptive*, dapat diketahui bahwa

pemodelan COD sungai di Surabaya yang terbaik diperoleh pada model GWR dengan fungsi kernel *Adaptive* Bisquare

## V. KESIMPULAN

Hasil pemodelan MGWR COD sungai di Surabaya dengan fungsi kernel *Fixed* Bisquare menghasilkan variabel prediktor global yang signifikan Nitrat, sedangkan variabel prediktor lokal yang signifikan adalah kecepatan aliran air dan Nitrit. Namun model MGWR tidak memberikan hasil yang berbeda dengan regresi global, sehingga model yang lebih baik untuk digunakan dalam memodelkan COD sungai di Surabaya adalah model GWR dengan fungsi kernel *Adaptive* Bisquare.

## DAFTAR PUSTAKA

- [1] D. Kusumawardani, *Valuasi Ekonomi Air Bersih di Surabaya (Studi Kasus Pada Air PDAM)*. Yogyakarta: Lembaga Penelitian dan Pengabdian Kepada Masyarakat UGM, (2010).
- [2] BBTkL-PPM. *Laporan Situasi dan Kecenderungan Parameter Pencemaran Air Badan Air Serta Risiko Gangguan Kesehatan Di Kali Surabaya Semester II*. Surabaya: Balai Besar Teknik Kesehatan Lingkungan dan Pemberantasan Penyakit Menular, (2010).
- [3] D.S.Utami, *Analisis Chemical Oxygen Demand (COD) pada Limbah Cair Domestik dengan Metode Spektrofotometri Portable*. Medan: Program D3 Analisis Farmasi dan Makanan Fakultas Farmasi Universitas Sumatera Utara, (2011).
- [4] C. Mei, N. Wang, dan W. Zhang, "Testing the importance of the explanatory variables in a mixed geographically weighted regression model". *Environment and Planning A*, (2006) 587-598.
- [5] H. Yasin, *Model Mixed Geographically Weighted Regression, Studi Kasus : Persentase Rumah Tangga Miskin di Kabupaten Mojokerto Tahun 2008*. Program Magister Jurusan Statistika FMIPA-ITS, (2011).
- [6] A.C. Rencher dan G. B.Schaalje, *Linear Models in Statistics Second Edition*. Hoboken: John Wiley and Sons Inc., (2008).
- [7] P. Propastin, M. Kappas, dan S. Erasmi, "Application of Geographically Weighted Regression to Investigate the Impact of Scale on Prediction Uncertainty by Modelling Relationship between Vegetation and Climate". *International Journal of Spatial Data Infrastructures Research Vol. 3*, (2008)73-94.
- [8] C. Brunson, S. Fotheringham, dan M. Charlton, *Geographically Weighted Regression as a Statistical Model*. Newcastle: Spatial Analysis Research Group Department of Geography University of Newcastle, (2000).
- [9] C.C.Yrigoyen, I.G. Rodríguez, dan J.V. Otero, *Modeling Spatial Variations in Household Disposable Income with Geographically Weighted Regression*. Madrid: Instituto L.R. Klein, (2006).
- [10] A.S. Fotheringham, M.E. Charlton, dan C. Brunson, "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis". *Environment and Planning A 1998*, Vol. 30, (1997) 1905-1927.
- [11] Puhadi dan H.Yasin, "Mixed Geographically Weighted Regression Model (Case Study: the Percentage of Poor Households in Mojokerto 2008)". *European Journal of Scientific Research ISSN 1450-216X Vol.69 No.2*, (2012) 188-196.
- [12] R. Koesnariyanto, *Pemodelan Indikator Pencemaran Air Secara Kimia (BOD) Dengan Geographically Weighted Regression*. Surabaya: Program Magister Fakultas Kesehatan Masyarakat Universitas Airlangga, (2012)..