

Model *Nonparametric* GWR untuk Identifikasi Faktor yang Mempengaruhi COD DAS Mahakam

CLEMENSUS ARLES¹, SIFRIYANI², FIDIA DENY TISNA AMIJAYA³

^{1,2}Program Studi Statistika, Jurusan Matematika, Fakultas MIPA, Universitas Mulawarman, Samarinda,
Indonesia

³Program Studi Matematika, Jurusan Matematika, Fakultas MIPA, Universitas Mulawarman, Samarinda,
Indonesia

e-mail: arlesclemen@gmail.com

ABSTRAK

Model Regresi *Spline* Nonparametrik dengan Pembobot Geografis merupakan pengembangan model regresi nonparametrik untuk data spasial dengan estimator parameter bersifat lokal untuk setiap pengamatan yang di aplikasikan pada data spasial. Data penelitian ini merupakan data sekunder yang diperoleh dari Dinas Lingkungan Hidup Provinsi Kalimantan Timur Samarinda. Tujuan penelitian ini adalah menentukan model pada data *Chemical Oxygen Demand* (COD) di Daerah Aliran Sungai Mahakam Kalimantan Timur dan untuk mengetahui faktor-faktor apa saja yang mempengaruhi COD di 28 titik pengambilan sampel pada DAS Mahakam tahun 2018. Pada Metode Regresi *Spline* Nonparametrik dengan Pembobot Geografis terdapat pemilihan titik knot optimum dengan menggunakan kriteria Generalized Cross-Validation (GCV) yang terkecil. Pembobot spasial yang digunakan adalah fungsi kernel bisquare. Hasil penelitian menunjukkan bahwa faktor-faktor yang berpengaruh terhadap COD di DAS Mahakam adalah Dissolved Oxygen (DO), Nitrit, dan tingkat keasaman pH.

Kata Kunci: *Chemical Oxygen Demand*, Pembobot Geografis, GCV, *Knot* dan *Spline*.

ABSTRACT

Nonparametric Spline Regression Model with Geographical Weights is the development of a nonparametric regression model for spatial data with a local parameter estimator for each observation that is applied to spatial data. This research data is secondary data obtained from the Environmental Service of the Province of East Kalimantan, Samarinda. The purpose of this study is to determine the model on Chemical Oxygen Demand (COD) data in the Mahakam River Basin, East Kalimantan and to determine what factors affect COD at 28 sampling points in the Mahakam Watershed in 2018. In the Nonparametric Spline Regression Method with Geographical Weighting, there is an optimum knot point selection using the smallest Generalized Cross-Validation (GCV) criteria. The spatial weighting used is the bisquare kernel function. The results showed that the factors that affect COD in the Mahakam watershed are Dissolved Oxygen (DO), Nitrite, and pH acidity.

Keywords: Chemical Oxygen Demand, Geographical Weighting, GCV, Knot and Spline.

1. PENDAHULUAN

Analisis regresi merupakan metode statistika yang banyak digunakan dalam berbagai bidang. Analisis tersebut bertujuan untuk mengetahui hubungan antara variabel respon dan variabel prediktor (Drapper dan Smith, 1992). Secara umum terdapat tiga pendekatan dalam menganalisis regresi yaitu pendekatan parametrik, semiparametrik dan nonparametrik. Jika pola data menunjukkan kecenderungan data mengikuti pola linear, kuadratik atau kubik maka menggunakan pendekatan regresi parametrik, pendekatan regresi semiparametrik digunakan apabila sebagian polanya diketahui dan sebagian tidak diketahui, namun apabila *scatterplot* tidak menunjukkan kecenderungan pola data tertentu maka pendekatan yang digunakan adalah regresi nonparametrik (Budiantara 2005 dan A, Islamiyati 2017).

Dalam penelitian ini penulis menggunakan pendekatan regresi nonparametrik karena tidak semua data lapangan diketahui bentuk pola datanya sehingga pendekatan ini mampu menyelesaikan permasalahan untuk pola data yang tidak diketahui, dan model yang dihasilkan dari model regresi nonparametrik sangat fleksibel dalam mendekati pola data. Pendekatan regresi nonparametrik nantinya membentuk estimasi sendiri tanpa dipengaruhi oleh subjektivitas peneliti. Terdapat beberapa estimasi dalam regresi nonparametrik, dalam penelitian ini penulis menggunakan estimasi *spline*. Salah satu kelebihan estimasi *Spline* adalah model cenderung mencari sendiri estimasi data kemanapun pola data tersebut bergerak. Kelebihan ini terjadi karena dalam *Spline* terdapat titik-titik knot, yaitu titik perpaduan bersama yang menunjukkan terjadinya perubahan pola perilaku data. *Spline* adalah salah satu jenis *piecewise* polinomial, yaitu polinomial yang memiliki sifat tersegmen. Sifat tersegmen ini memberikan fleksibilitas lebih dari polinomial biasa, sehingga memungkinkan untuk menyesuaikan diri secara lebih efektif terhadap karakteristik lokal suatu fungsi atau data (Budiantara, 2009).

Model regresi nonparametrik *spline truncated* dalam penelitian ini diaplikasikan pada data COD sebagai parameter kualitas air sungai mahakam pada 28 titik pengambilan sampel di DAS Mahakam Provinsi Kalimantan Timur, COD menggambarkan bahan organik yang mudah terurai maupun sukar terurai. Secara umum parameter COD digunakan untuk mengukur pencemaran limbah rumah tangga, Aktivitas pertambangan maupun aktivitas kapal yang membuang limbah mengakibatkan menurunnya jumlah oksigen dalam air sehingga dapat merusak ekosistem air sungai. Pemilihan model *spline* terbaik dapat dilihat dari nilai knot yang optimum berdasarkan nilai GCV yang terkecil.

Berdasarkan hal tersebut penulis ingin melakukan penelitian tentang regresi nonparametrik *spline truncated* untuk memodelkan COD di DAS Mahakam dengan menggunakan tiga variabel prediktor yaitu DO, Nitrit dan Tingkat Keasaman (pH).

2. METODE PENELITIAN

Data yang digunakan dalam penelitian ini menggunakan data sekunder dari Dinas Lingkungan Hidup (DLH) Provinsi Kalimantan Timur. Di 28 titik pengambilan sampel pada tahun 2018. Variabel respon yang digunakan adalah COD (y), DO (x_1), Nitrit (x_2) dan pH (x_3).

Berikut langkah-langkah untuk analisis regresi *spline truncated* dalam penelitian ini.

1. Identifikasi variabel respon dan variabel prediktor .
2. Analisis pola linier antara variabel respon dengan masing-masing variabel prediktor.
3. Menghitung jarak Euclidean antara lokasi ke- i yang terletak pada koordinat (u_i, v_i) terhadap lokasi ke- j yang terletak pada koordinat (u_j, v_j)
4. Menentukan pembobot untuk model regresi nonparametrik pembobot fungsi kernel bisquare
5. Menentukan titik knot optimum dengan menggunakan metode GCV
6. Mendapatkan estimasi model regresi nonparametrik pembobot geografis
7. Pengujian hipotesis simultan untuk parameter model regresi nonparametrik pembobot geografis
8. Pengujian hipotesis parsial atau pengujian pengaruh lokasi secara parsial pada setiap variabel prediktor untuk model regresi nonparametrik pembobot geografis menggunakan
9. Interpretasi model regresi nonparametrik pembobot geografis dan menarik kesimpulan.

3. HASIL DAN PEMBAHASAN

Analisis Statistika Deskriptif

Pada table 1 berikut ini diberikan statistika deskriptif dari data penelitian yang digunakan. Statistika deskriptif yang akan ditampilkan adalah nilai minimum, maximum dan ukuran pemusatan.

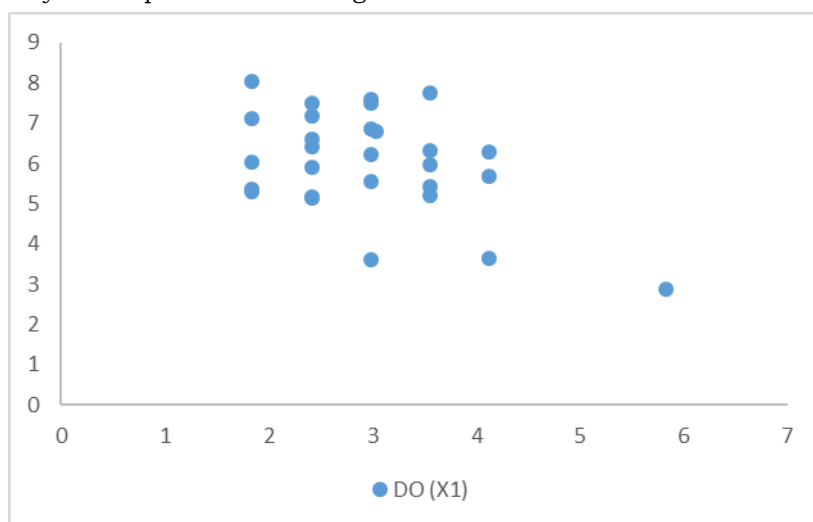
Tabel 1 Statistika Deskriptif Variabel Respon dan Variabel Prediktor

Variabel	Data	Minimum	Maksimum	Rata-rata
COD	28	1,84	5,83	2,9610
DO	28	2,88	8,05	6,0390
Nitrit	28	0,22	2,25	0,9010
pH	28	6,00	7,88	7,0661

Berdasarkan table 1 dapat diketahui bahwa nilai rata-rata nilai COD di sungai Mahakam sebesar 2,9610 mg/l. sedangkan rata-rata kadar DO sebesar 6,0390 mg/l, rata-rata kadar nitrit sebesar 0,9010 mg/l dan rata-rata pH sebesar 7,0661.

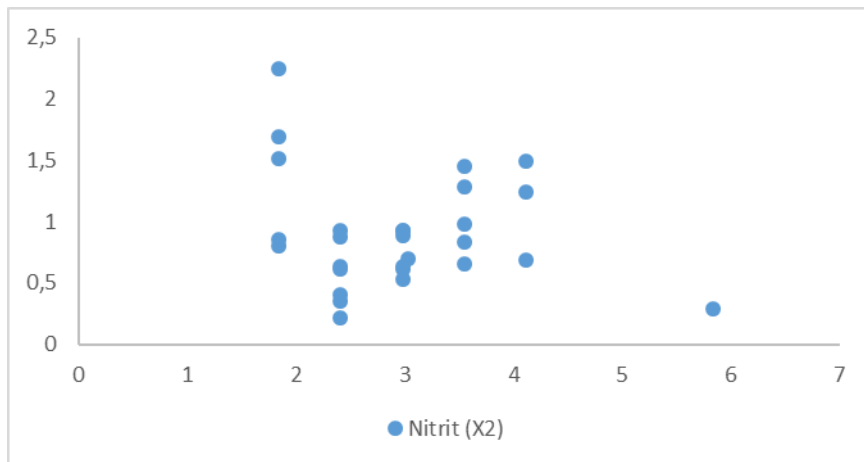
Identifikasi pola data dengan *scatterplot*

Scatterplot dibuat untuk mengetahui pola data antara dua variabel. Dalam analisis ini ingin diketahui pola data antara variabel respon terhadap variabel prediktor. *Scatterplot* dilakukan secara satu persatu dengan masing-masing variabel prediktor sehingga *scatterplot* yang terbentuk sebanyak 3 dapat dilihat dalam gambar berikut.



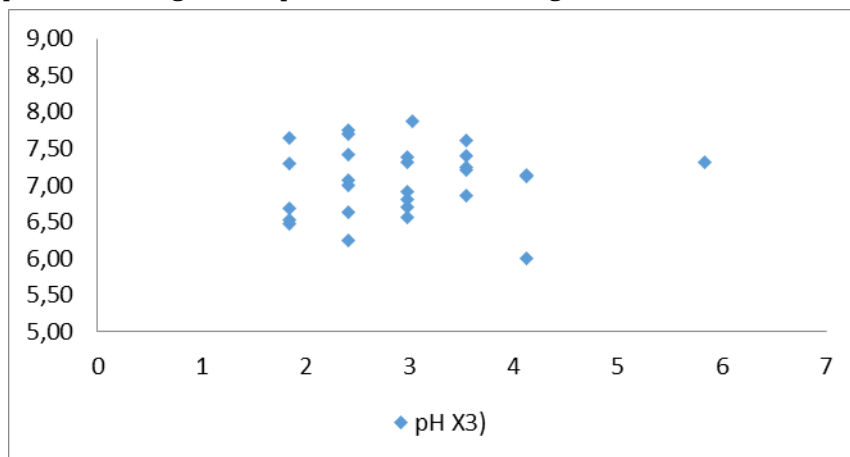
Gambar 1 *scatterplot* COD terhadap DO

Dari Gambar 1 terlihat bahwa pola hubungan antara COD dan DO memiliki pola yang berubah-ubah pada sub-sub interval tertentu. Sehingga tidak menunjukkan adanya kecenderungan pola atau pola tidak diketahui bentuk kurva regresinya, Karena pola data tidak diketahui bentuk regresinya, maka untuk mengestimasi akan digunakan regresi nonparametrik.



Gambar 2 scatterplot COD terhadap Nitrit

Dari Gambar 2 terlihat bahwa pola hubungan antara COD dan Nitrit tidak menunjukkan adanya kecenderungan pola atau pola tidak diketahui bentuk kurva regresinya. Sehingga dapat digunakan pendekatan regresi nonparametrik dalam mengestimasi.



Gambar 3 scatterplot COD terhadap pH

Dari Gambar 3 5 terlihat bahwa pola hubungan antara COD dan pH tidak menunjukkan adanya kecenderungan pola atau pola tidak diketahui bentuk kurva regresinya. Sehingga dapat digunakan pendekatan regresi nonparametrik dalam mengestimasi.

Pemilihan Titik Knot Optimal

Pemilihan model *spline truncated* terbaik diperoleh berdasarkan nilai GCV minimum dan nilai R^2 yang maksimum. Berikut ditampilkan nilai GCV dan R^2 untuk 2 dan 3 titik knot yang disajikan dalam tabel 2.

Tabel 2 Nilai GCV Masing-masing Titik Knot

Titik Knot	GCV Minimum	R^2
2 Titik knot	0,4306	76,8400
3 Titik knot	0,6058	54,5150

Dari dua nilai GCV diperoleh nilai GCV yang paling minimum dan nilai R^2 paling maksimum adalah pada 2 titik knot.

Model Regresi Spline Truncated Optimal

Berdasarkan titik knot optimal yang terpilih yaitu 2 titik knot, sehingga dapat dibentuk model regresi nonparametrik *spline truncated* sebagai berikut:

$$\hat{y}_{23} = 593,6385 - 17,9433x_{1,23} - 3,2491x_{1,23}^2 - 2,5332(x_{1,23} - 4,4827)_+^2 + 2,2366(x_{1,23} - 6,8092)_+^2 - 172,7011x_{2,23} + 2,1264x_{2,23}^2 - 4,4644(x_{2,23} - 0,8493)_+^2 - 1,7735(x_{2,23} - 1,7628)_+^2 + 2,9354x_{3,23} + 13,4595x_{3,23}^2 - 18,5597(x_{3,23} - 6,5828)_+^2 + 18,6099(x_{3,23} - 7,5288)_+^2$$

Interpretasi Model

1. Apabila variabel nitrit dan tingkat keasaman pH dianggap konstan maka model COD (Y) terhadap konsentrasi DO (X_1) adalah sebagai berikut

$$\hat{y}_{23} = 593,6385 - 17,9433x_{1,23} - 3,2491x_{1,23}^2 - 2,5332(x_{1,23} - 4,4827)_+^2 + 2,2366(x_{1,23} - 6,8092)_+^2$$

$$= \begin{cases} 593,6385 - 17,9433x_{1,23} - 3,2491x_{1,23}^2 & x_{1,23} < 4,4827 \\ 644,5421 - 40,6545x_{1,23} - 0,7159x_{1,23}^2 & 4,4827 \leq x_{1,23} \leq 6,8092 \\ 748,2425 - 71,1134x_{1,23} - 1,5207x_{1,23}^2 & x_{1,23} \geq 6,8092 \end{cases}$$

berdasarkan model dijelaskan DO (X_1) kurang dari 4,4827 mg/l yaitu sebesar 3,61 mg/l maka setiap kenaikan 1 mg/l COD mengakibatkan penurunan konsentrasi DO sebesar 21,1924 mg/l.

2. Apabila variabel Do dan tingkat keasaman pH dianggap konstan maka model COD (Y) terhadap Nitrit (X_2) adalah sebagai berikut sebagai berikut

$$\hat{y}_{23} = 593,6385 - 172,7011x_{2,23} + 2,1264x_{2,23}^2 - 4,4644(x_{2,23} - 0,8493)_+^2 - 1,7735(x_{2,23} - 1,7628)_+^2$$

$$= \begin{cases} 593,6384 - 172,7011x_{2,23} + 2,1264x_{2,23}^2 & x_{2,23} < 0,8493 \\ 590,4183 - 165,1179x_{2,23} - 2,3380x_{2,23}^2 & 0,8493 \leq x_{2,23} \leq 1,7628 \\ 595,9295 - 171,3706x_{2,23} - 0,5645x_{2,23}^2 & x_{2,23} \geq 1,7628 \end{cases}$$

berdasarkan model dijelaskan Nitrit (X_2) terletak di antara 0,8493 mg/l dan 1,7628 Mg/l maka setiap kenaikan 1 mg/l COD (Y) akan mengakibatkan penurunan Nitrit (X_2) sebesar 167,4559 mg/l.

3. Apabila variabel DO, dan Nitrit dianggap konstan maka model COD (Y) terhadap tingkat keasaman pH (X_3) adalah sebagai berikut

$$\hat{y}_{23} = 593,6385 + 2,9354x_{3,23} + 13,4595x_{3,23}^2 - 18,5597(x_{3,23} - 6,5828)_+^2 + 18,6099(x_{3,23} - 7,5288)_+^2$$

$$= \begin{cases} 593,6384 - 2,9354x_{3,23} + 13,4595x_{3,23}^2 & x_{3,23} < 6,5828 \\ 210,6145 - 10,2302x_{3,23} - 5,1002x_{3,23}^2 & 6,5828 \leq x_{3,23} \leq 7,5288 \\ 1265,4757 - 290,4506x_{3,23} + 13,5097x_{3,23}^2 & x_{3,23} \geq 7,5288 \end{cases}$$

berdasarkan model dijelaskan tingkat keasaman $\text{pH}(X_3)$ berada diantara 6,5828 dan 7,5288 maka setiap penambahan 1 pH akan mengakibatkan penurunan COD sebanyak 15,3304 Mg/l.

4. SIMPULAN DAN SARAN

Berdasarkan hasil pengujian yang telah dilakukan maka kesimpulan yang dapat di ambil adalah sebagai berikut:

1. Setiap kenaikan 1 Mg/l konsentrasi COD akan mengakibatkan penurunan konsentrasi DO sebesar 21,1924 Mg/l, setiap kenaikan 1 Mg/l COD akan mengakibatkan penurunan Nitrit sebesar 167,4559 Mg/l, dan setiap penambahan 1 pH akan mengakibatkan penurunan COD sebanyak 15,3304 Mg/l.
2. Salah satu model regresi nonparametrik dengan pembobot geografis pada titik lokasi AS. KRM. Yaitu

$$\begin{aligned} \hat{y}_{23} = & 593,6385 - 17,9433x_{1,23} - 3,2491x_{1,23}^2 - 2,5332(x_{1,23} - 4,4827)_+^2 + \\ & 2,2366(x_{1,23} - 6,8092)_+^2 - 172,7011x_{2,23} + 2,1264x_{2,23}^2 - \\ & 4,4644(x_{2,23} - 0,8493)_+^2 - 1,7735(x_{2,23} - 1,7628)_+^2 + \\ & 2,9354x_{3,23} + 13,4595x_{3,23}^2 - 18,5597(x_{3,23} - 6,5828)_+^2 + \\ & 18,6099(x_{3,23} - 7,5288)_+^2 \end{aligned}$$

3. Data yang digunakan pada penelitian ini adalah *Chemical Oxygen Demand (COD)* DAS Mahakam pada tahun 2018, dengan variabel-variabel yang mempengaruhi adalah *Dissolved Oxygen (DO)*, Nitrit, dan Tingkat Keasaman (pH)

Saran

Pada penelitian ini digunakan fungsi pembobot yaitu fungsi kernel Bisquare, sehingga untuk penelitian selanjutnya dapat menggunakan pembobot fungsi kernel Gaussian dan fungsi kernel Tricube.

DAFTAR PUSTAKA

- Budiantara, I Nyoman. (2005). Model Spline Multivariabel dalam regresi nonparametrik, *Makalah Seminar Nasional Matematika, Jurusan Matematika FMIPA*. ITS. Surabaya.
- Budiantara, I Nyoman. (2009). *Spline dalam Regresi Nonparametrik: Sebuah Pemodelan Statistika Masa Kini dan Masa Mendatang*. Surabaya: FMIPA ITS.
- Draper dan Smith. (1992). *Applied Regression Analysis*, Second Edition. New York: John Wiley & Sons.
- Islamiyati, Anna. 2017. *Spline Polynomial* Jurnal Matematika, Statistika dan Komputasi. Vol 14, No.1, 54-60, Juli 2017. E-ISSN: 2614-8811. Jurnal Teknologi Pertanian Andalas. Vol. 23, No.1, Maret 2019 ISSN 1410-1920.