

Pemodelan Konsentrasi Partikel Debu (PM₁₀) pada Pencemaran Udara di Kota Surabaya dengan Metode *Geographically-Temporally Weighted Regression*

Kurniasari Aisyiah, Sutikno, dan I Nyoman Latra

Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Teknologi Sepuluh Nopember (ITS)

Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia

e-mail: sutikno@statistika.its.ac.id

Abstrak - Konsentrasi partikel debu (PM₁₀) di Kota Surabaya menempati urutan pertama di Jawa Timur. Hal ini karena aktifitas penduduk Kota Surabaya yang tinggi menyebabkan polusi udara. Partikel debu (PM₁₀) merupakan salah satu polutan yang apabila terhisap langsung ke dalam paru-paru dan mengendap di alveoli dapat membahayakan sistem pernafasan. Dalam pemantauan kualitas udara, seringkali peralatan pengukur konsentrasi PM₁₀ mengalami kerusakan, sehingga data polutan tersebut tidak terukur atau tidak tersedia (missing), maka perlu dilakukan pendugaan data PM₁₀ pada lokasi yang tidak terukur. Salah satu metode yang digunakan adalah *Geographically-Temporally Weighted Regression (GTWR)* untuk memprediksi konsentrasi PM₁₀ dengan menggunakan parameter meteorologi. Konsentrasi partikel debu bergantung pada lokasi dan waktu. Hasil penelitian menyimpulkan bahwa kondisi pencemaran udara di Kota Surabaya pada tahun 2010 masih dinyatakan baik dan metode GTWR memberikan hasil yang lebih akurat daripada regresi nonspasial karena dapat mengakomodasi adanya pengaruh heterogenitas spasial dan temporal pada konsentrasi partikel debu (PM₁₀).

Kata Kunci — Partikel debu (PM₁₀), Regresi, Spasial, Temporal

I. PENDAHULUAN

Adanya peningkatan pada komunikasi, inovasi, dan transportasi merupakan dampak positif globalisasi. Namun, secara bersamaan globalisasi turut memberikan dampak pada siklus ekologis berupa polusi. Salah satu polusi pada lingkungan hidup adalah polusi udara. Berdasarkan Baku Mutu Udara Ambien Nasional terdapat 9 jenis polutan dengan nilai baku mutu yakni Sulfur Dioksida (SO₂) 0,1 ppm tiap pengukuran 24 jam, Karbon Monoksida (CO) 20 ppm tiap pengukuran 24 jam, Nitrogen Dioksida (NO₂) 0,05 ppm tiap pengukuran 24 jam, Oksidan (O₃) 0,1 ppm tiap pengukuran 1 jam, partikel ≤ 10 μm (PM₁₀) 150 mg/m³ tiap pengukuran 24 jam, partikel ≤ 2,5 μm (PM_{2,5}) 65 ug/Nm³ tiap pengukuran 24 jam, TSP 230 ug/Nm³ tiap pengukuran 24 jam, Timah Hitam (Pb) 2 ug/Nm³ tiap pengukuran 24 jam, dan Debu Jatuh 10 ton/km²/30hari. Jika kadarnya melebihi nilai baku mutu maka udara ambien dinyatakan tercemar.

Salah satu polutan yang menyebabkan polusi adalah partikel debu (PM₁₀). Dengan uji toksikologi dapat memberikan hasil bahwa partikel debu (PM₁₀) yang terhisap langsung ke dalam paru-paru dan mengendap di alveoli dapat membahayakan sistem [1]. Kadar partikel debu

(PM₁₀) di Kota Surabaya menempati urutan tertinggi di Jawa Timur [2]. Dibandingkan dengan kadar O₃, SO₂, dan CO, partikel debu (PM₁₀) memiliki konsentrasi maksimum yang lebih tinggi di area Kota Surabaya [3]. Peralatan yang digunakan untuk mengukur adalah *Hi-Vol* dengan metode *Gravimetric*. Seringkali peralatan tersebut mengalami kerusakan sehingga konsentrasi partikel debu (PM₁₀) tidak dapat diketahui. Namun, mesin pada Unit Stasiun Pemantau Kualitas Udara (SUF) masih dapat mengukur parameter meteorologi seperti suhu, kelembaban, kecepatan angin dan arah angin. Untuk itu diperlukan pemodelan untuk memprediksi konsentrasi partikel debu (PM₁₀) dengan menggunakan parameter meteorologi.

Penelitian yang membahas pencemaran udara di Kota Surabaya dilakukan oleh Putri yang mengestimasi Nitrogen Dioksida (NO₂) dan Karbon Monoksida (CO) dengan metode interpolasi cokigring[4]. Sementara Choiruddin meneliti kadar BOD Kali Surabaya menggunakan *Geographically-Temporally Weighted Regression (GTWR)* dan menyimpulkan bahwa terdapat efek heterogen spasial dan temporal [5]. Selain itu Huang menggunakan metode GTWR untuk memodelkan variasi harga rumah di Calgary, Canada [6]. Pada penelitian ini metode GTWR digunakan untuk memodelkan partikel debu dengan mengakomodasi adanya pengaruh heterogenitas spasial dan temporal pada konsentrasi partikel debu (PM₁₀).

II. TINJAUAN PUSTAKA

A. Regresi Linier

Analisis regresi linier merupakan suatu metode untuk menjelaskan hubungan antara variabel respon dan variabel prediktor secara linier [7]. Persamaan umum untuk model regresi linier adalah sebagai berikut.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \dots + \beta_p X_{ip} + \varepsilon_i$$

Dalam bentuk matriks, persamaan umum model regresi linier adalah sebagai berikut.

$$Y = X\beta + \varepsilon$$

dengan,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}; X = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots & X_{1p} \\ 1 & X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}; \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix};$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}; \begin{matrix} n = \text{jumlah data} \\ p = \text{jumlah prediktor} \end{matrix}$$

Dengan menggunakan metode *Ordinary Least Square* (OLS) untuk meminimumkan jumlah kuadrat error, maka didapatkan estimator parameter β yakni $\hat{\beta} = (X^T X)^{-1} X^T Y$.

1. *Pengujian Hipotesis Model Regresi*

Pengujian hipotesis secara serentak dilakukan untuk mengetahui kesesuaian model. Tabel 1 menjelaskan nilai perhitungan pada analisis varians pada model regresi.

Tabel 1. Analisis Varians

Sumber Variasi	Jumlah Kuadrat (SS)	Derajat Bebas (DF)	Rata-rata Kuadrat (MS)	F
Regresi	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	p	$\frac{SSR}{p}$	$\frac{MSR}{MSE}$
Error	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$n - (p + 1)$	$\frac{SSE}{n - (p + 1)}$	
Total	$\sum_{i=1}^n (y_i - \bar{y})^2$	$n - 1$		

Hipotesis pada pengujian serentak adalah sebagai berikut.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal ada satu } \beta_k \neq 0 ; k = 1, 2, \dots, p$$

Statistik uji yang diperoleh melalui perhitungan nilai F yang tercantum pada Tabel 2.1 dan daerah kritis yakni tolak H_0 apabila $F > F_{\alpha;p;(n-p-1)}$ atau $p\text{-value} < \alpha$. Jika diperoleh keputusan tolak H_0 , maka dapat disimpulkan bahwa terdapat minimal satu parameter β yang signifikan terhadap respon. Kemudian dilakukan pengujian secara parsial untuk mengetahui parameter β yang signifikan terhadap respon dengan hipotesis sebagai berikut.

$$H_0 : \beta_k = 0$$

$$H_1 : \beta_k \neq 0 ; k = 1, 2, \dots, p$$

Statistik uji yang diperoleh melalui perhitungan nilai t sebagai berikut.

$$t = \frac{b_k}{s} (\sum_{i=1}^n (x_i - \bar{x})^2)^{\frac{1}{2}} = \frac{\hat{\beta}_k}{SE(\hat{\beta}_k)}$$

Daerah kritis pada pengujian secara parsial yakni tolak H_0 apabila $|t| > t_{\frac{\alpha}{2};(n-p-1)}$ atau $p\text{-value} < \alpha$. Jika diperoleh keputusan tolak H_0 , maka dapat disimpulkan bahwa β_k signifikan terhadap respon.

2. *Pengujian Asumsi Residual Model Regresi*

Pada analisis regresi terdapat beberapa asumsi residual yang harus dipenuhi yakni residual berdistribusi normal, identik, dan independen. Pengujian asumsi distribusi normal menggunakan uji *Kolmogorov-Smirnov*. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : F_n(e) = F_0(e)$$

(residual mengikuti distribusi normal)

$$H_1 : F_n(e) \neq F_0(e)$$

(residual tidak mengikuti distribusi normal)

Statistik uji yang diperoleh melalui perhitungan nilai $D = \sup |F_n(e) - F_0(e)|$ dan daerah kritis yakni tolak H_0 apabila $|D| > D_{\alpha;n}$ atau $p\text{-value} < \alpha$.

Untuk pengujian asumsi identik menggunakan uji *Glejser*. Hipotesis yang digunakan adalah sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (residual bersifat identik)}$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, n$$

(residual tidak bersifat identik)

Statistik uji yang yang digunakan, diperoleh dari perhitungan nilai $F = \frac{(\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}|)^2)/(p)}{(\sum_{i=1}^n (|\hat{e}_i| - |\bar{e}|)^2)/(n-p-1)}$. Pengujian akan memberikan keputusan tolak H_0 apabila $F > F_{\alpha;p;(n-p-1)}$ atau $p\text{-value} < \alpha$.

Untuk pengujian asumsi independen menggunakan uji *Durbin-Watson*. Hipotesis yang digunakan sebagai berikut.

$$H_0 : \rho = 0 \text{ (residual bersifat independen)}$$

$$H_1 : \rho \neq 0 \text{ (residual tidak bersifat independen)}$$

Statistik uji yang diperoleh melalui perhitungan nilai d yakni sebagai berikut.

$$d = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Pengujian akan memberikan keputusan tolak H_0 apabila nilai $d < d_{L;\alpha}$ atau $d > 4d_{U;\alpha}$

Adanya multikolinieritas yakni terdapat korelasi antar variabel prediktor menyebabkan kesalahan pada pendugaan parameter dalam pemodelan regresi linier. Untuk mengetahui adanya multikolinieritas dilakukan dengan uji *Variance Inflation Factor* (VIF). Nilai *VIF* didapatkan melalui perhitungan sebagai berikut.

$$VIF_j = \frac{1}{1 - R^2_j}$$

dengan R^2_j adalah koefisien determinasi antara variabel prediktor X_j dan variabel prediktor yang lain. Apabila nilai *VIF* lebih besar dari 10, maka dinyatakan adanya multikolinieritas.

B. *Uji Heterogenitas*

Untuk mengetahui adanya heterogenitas spasial maka dilakukan pengujian menggunakan uji *Breusch-Pagan*. Hipotesis yang digunakan pada uji *Breusch-Pagan* adalah sebagai berikut.

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2 = \sigma^2 \text{ (homoskedasitisitas)}$$

$$H_1 : \text{minimal ada satu } \sigma_i^2 \neq \sigma^2 ; i = 1, 2, \dots, n$$

(heteroskedasitisitas)

Statistik uji yang diperoleh melalui perhitungan nilai *BP* sebagai berikut.

$$BP = \frac{1}{2} f^T Z (Z^T Z)^{-1} Z^T f$$

yang mana nilai vektor f adalah $f_i = \frac{e_i^2}{\sigma^2} - 1$. Nilai e_i adalah *least square residual* untuk observasi ke- i dan Z merupakan matriks berukuran $n \times (p + 1)$ dengan elemen vektor yang sudah dinormal-standarkan untuk setiap observasi. Pengujian akan memberikan keputusan tolak H_0 apabila $BP > \chi^2_{\alpha;(p+1)}$.

Untuk mengetahui adanya heterogenitas temporal maka dilakukan dengan menggambarkan data menggunakan *boxplot*. *Boxplot* digunakan untuk membandingkan karakter distribusi nilai data secara individual atau kelompok kategori dari suatu variabel serta untuk menangkap adanya data yang *outlier*.

C. *Geographically-Temporally Weighted Regression*

Geographically-Temporally Weighted Regression (GTWR) merupakan metode pengembangan dari *Geographically Weighted Regression* yang mengakomodasi adanya heterogenitas secara spasial (lokasi) dan secara

temporal (waktu) [8]. Persamaan umum untuk model GTWR adalah sebagai berikut.

$$Y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i) X_{ik} + \varepsilon_i$$

Pada model p adalah jumlah variabel prediktor dan i menunjukkan observasi. Estimasi nilai parameter didapatkan dari perhitungan sebagai berikut.

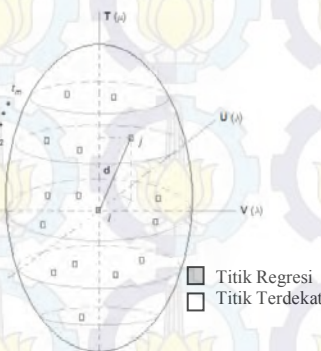
$$\hat{\beta}(u_i, v_i, t_i) = (X^T W(u_i, v_i, t_i) X)^{-1} X^T W(u_i, v_i, t_i) Y$$

yang mana $W(u_i, v_i, t_i) = \text{diag}(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$ dengan n adalah jumlah data observasi. Elemen diagonal α_{ij} ($1 \leq j \leq n$) adalah fungsi jarak ruang (u, v) dan waktu dari (t) sesuai dengan pembobotan pada regresi yang berdekatan dengan titik pengamatan i . Semakin dekat titik yang diamati dengan titik i , maka koordinat memiliki pengaruh yang lebih besar pada estimasi parameter.

Besar pembobotan ditentukan menggunakan fungsi *kernel gaussian* karena menghasilkan hasil yang lebih halus dan standar error yang lebih kecil pada estimasi parameter [9]. Fungsi jarak berdasarkan fungsi *kernel gaussian* adalah sebagai berikut.

$$W_{ij} = \exp\left(-\frac{(d_{ij}^{ST})^2}{h^2_{ST}}\right)$$

dengan d_{ij} merupakan jarak antara titik i dan titik j yang diperoleh dari fungsi jarak *eucclidean* yakni $(d_{ij}^S)^2 = (u_i^2 - u_j^2) + (v_i^2 - v_j^2)$ dan $(d_{ij}^T)^2 = (t_i^2 - t_j^2)$. Dan h adalah parameter non negatif untuk penghalus atau biasa disebut *bandwith*. Adanya perbedaan skala secara spasial dan temporal maka sistem koordinat yang digunakan adalah *ellipsoidal*.



Gambar 1 Ilustrasi Jarak Spasial-Temporal (Sumber : Huang, dkk, 2010)

Pada Gambar 1 dijelaskan bahwa fungsi jarak spasial-temporal dibentuk melalui kombinasi fungsi jarak spasial (d^S) dan fungsi jarak temporal (d^T). Maka fungsi jarak spasial-temporal adalah sebagai berikut.

$$(d^{ST})^2 = \lambda(d^S)^2 + \mu(d^T)^2 \tag{1}$$

dengan λ dan μ menyatakan faktor skala penyeimbang perbedaan efek yang digunakan untuk mengukur jarak spasial dan temporal. Menurut Huang dengan mensubstitusikan fungsi jarak *eucclidean* maka persamaan (1) menjadi sebagai berikut.

$$(d^{ST}_{ij})^2 = \lambda[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2 \tag{2}$$

Kemudian dengan mensubstitusikan persamaan (2) pada fungsi jarak *kernel gaussian* maka didapatkan perhitungan sebagai berikut.

$$\begin{aligned} \alpha_{ij} &= \exp\left(-\frac{(d_{ij}^{ST})^2}{h^2_{ST}}\right) \\ &= \exp\left\{-\left(\frac{\lambda[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu(t_i - t_j)^2}{h^2_{ST}}\right)\right\} \end{aligned}$$

dengan h^2_{ST} adalah parameter *bandwith* spasial temporal dengan $h^2_S = h^2_{ST}/\lambda$ merupakan parameter *bandwith* spasial dan $h^2_T = h^2_{ST}/\mu$ merupakan parameter *bandwith* temporal. Dimisalkan τ merupakan parameter rasio μ/λ dengan $\lambda \neq 0$, maka persamaan (2) dapat dibagi dengan λ untuk memunculkan parameter τ sehingga dapat ditulis dalam bentuk sebagai berikut.

$$\frac{(d^{ST}_{ij})^2}{\lambda} = (u_i - u_j)^2 + (v_i - v_j)^2 + \tau(t_i - t_j)^2$$

Parameter τ didapatkan melalui metode optimasi koefisien determinasi (R^2) secara iteratif. Sehingga estimasi parameter τ dapat menghasilkan R^2 yang maksimum. Parameter τ digunakan untuk memperbesar atau memperkecil efek jarak temporal terhadap efek jarak spasial. Kemudian estimasi parameter μ dan λ didapatkan melalui metode iteratif berdasarkan estimasi parameter τ yang menghasilkan nilai R^2 maksimum. Untuk penentuan *bandwith* spasial temporal dapat diinisiasi dengan menggunakan *bandwith* spasial ditentukan oleh peneliti dengan *trial-error*.

D. Kriteria Keabakan Model

Kriteria kebaikan model yang digunakan adalah kriteria koefisien determinasi (R^2) dan *Mean Square Error (MSE)*. Kriteria R^2 merupakan metode menemukan himpunan variabel prediktor terbaik dalam memprediksi variabel respon melalui model regresi. Model terbaik ditunjukkan dengan nilai R^2 yang paling tinggi untuk setiap unit variabel prediktor dipertimbangkan dalam model. Formulasi perhitungan koefisien determinasi (R^2) adalah sebagai berikut.

$$R^2 = 1 - \frac{SSE}{SST}$$

Sementara *Mean Square Error (MSE)* digunakan untuk mengevaluasi tingkat kesalahan berdasarkan nilai residual pada model yang satu dan dibandingkan dengan model yang lainnya.

E. Pencemaran Udara

Udara dinyatakan tercemar apabila mengandung polutan yang kadarnya melebihi nilai baku mutu. Menurut Peraturan Pemerintah Republik Indonesia Nomor 41 Tahun 1999, polusi atau dapat disebut pencemaran udara adalah masuknya atau dimasukkannya zat, energi, dan/atau komponen lain ke dalam udara ambien oleh kegiatan manusia, sehingga mutu udara ambien turun sampai ke tingkat tertentu. Sedangkan udara ambien adalah udara bebas dipermukaan bumi pada lapisan troposfir yang berada di dalam wilayah yurisdiksi Republik Indonesia yang dibutuhkan dan mempengaruhi kesehatan manusia, makhluk hidup dan unsur lingkungan hidup lainnya.

Partikel debu (PM_{10}) merupakan salah satu jenis polutan dengan konsentrasi paling tinggi dibandingkan konsentrasi O_3 , SO_2 , dan CO di Kota Surabaya. Zusana menyebutkan bahwa faktor yang mempengaruhi konsentrasi partikel debu (PM_{10}) adalah suhu, kelembaban, kecepatan angin, dan arah angin. Artinya perbedaan tempat atau adanya heterogenitas

