KINERJA REGRESI PROSES GAUSSIAN UNTUK PEMODELAN KALIBRASI PEUBAH GANDA PADA DAERAH IDENTIFIKASI SPEKTRA INFRA MERAH SENYAWA AKTIF

Moch. Abdul Mukid[1], Aji Hamim Wigena[2], Erfiani[2]

mamukid@yahoo.com

Abstract

Multivariate calibration models have been developed usually by using principal component regression and partial least squares regression. This research proposes the application of Gaussian process regression as an alternative method to develop a calibration model. This method is applied to the measurement of curcumin concentration based on FTIR spectra. FTIR spectra was choosen at the identification are of curcuminoid. To handle the high dimensionality of spectra data, principal component analysis was initially performed, followed by applying the Gaussian process regression. Using three principal components, 99,66% of the original data's variability can be explained. This model was attempted for various covariance functions. The results indicate that the most relevant and suitable covariance function for curcumin concentration measurement was Matern 3 isotropic. The hyperparameter values for Matern 3 isotropic were estimated by *Maximum Marginal Likelihood* Method. Based on RMSEP criteria, the performance of Gaussian process regression with covariance function Matern 3 isotropic at infra red spectra indentification area is better than at wavelength number 4000-400 cm^{-1}.

Keywords: Gaussian Process Regression, Covariance Function, Hyperparameter, Maximum Marginal Likelihood Function.

Abstrak

Model-model kalibrasi peubah ganda biasanya dikembangkan dengan pendekatan regresi komponen utama dan regresi kuadrat terkecil parsial. Dalam penelitian ini diajukan penerapan regresi proses Gaussian sebagai sebuah metode alternatif untuk membangun sebuah model kalibrasi. Metode ini diterapkan pada pengukuran konsentrasi kurkumin berdasarkan atas data spektra FTIR. Data spektra FTIR dipilih hanya pada daerah identifikasi dari kurkuminoid. Untuk mengatasi dimensi dari data spektra, analisis komponen utama digunakan terlebih dahulu untuk mereduksi jumlah dimensinya. Diketahui bahwa dengan menggunakan tiga komponen utama 99,66% keragaman dari data asal dapat dijelaskan. Pada model ini dicoba penggunaan beberapa fungsi peragam dan diketahui bahwa fungsi peragam yang relevan adalah Matern 3 isotropik karena menghasilkan nilai RMSEP terkecil yaitu 0,5179. Berdasarkan kriteria atas nilai RMSEP, kinerja regresi proses Gaussian dengan fungsi peragam Matern 3 isotropik pada daerah identifikasi spektra infra merah lebih baik jika dibandingkan pada pemilihan bilangan gelombang 4000-400 cm^{-1}.

Kata kunci: Regresi Proses Gaussian, Fungsi Peragam, Hiperparameter, Fungsi Kemungkinan Marginal Maksimum.

[1] Dosen pada Program Studi Statistika Universitas Diponegoro Semarang
[2] Dosen pada Program Studi Statistika, Sekolah Pascasarjana, Institut Pertanian Bogor

Latar Belakang

Di Indonesia tanaman obat telah lama digunakan oleh masyarakat dan industri dalam pembuatan jamu. Penggunaan tanaman obat yang semakin meluas sudah selayaknya diikuti dengan usaha untuk menjamin kualitas tanaman obat tersebut. Hal ini untuk menjamin agar produksinya dapat bersaing dan diterima oleh masyarakat. Salah satu indikator kualitas tanaman obat adalah konsentrasi senyawa aktifnya.

Proses penentuan konsentrasi senyawa aktif yang dikandung oleh suatu tanaman obat perlu dilakukan secara cepat dan akurat. Secara kuantitatif dan kualitatif suatu senyawa aktif dapat diketahui antara lain melalui metode HPLC (*High Performance Liquid Chromatography*) dan FTIR (*Fourier Trasform Infrared*). Penentuan konsentrasi senyawa aktif dilakukan melalui proses yang panjang meliputi penghancuran bahan, pelarutan, dan pengukuran dengan HPLC. Proses ini memerlukan waktu dan biaya yang relatif mahal. Untuk itu sangat diperlukan metode yang handal tetapi relatif mudah untuk digunakan. Salah satu metodenya adalah dengan membuat sebuah model kalibrasi. Model ini menyatakan hubungan antara konsentrasi senyawa aktif hasil pengukuran HPLC dengan persen transmitan (absorban) yang diukur dengan menggunakan FTIR. Tujuan dari pembentukan model ini adalah untuk memprediksi konsentrasi senyawa aktif dengan akurasi yang tinggi dari nilai persen transmitan yang secara ekonomi lebih murah dan mudah diperoleh (Erfiani, 2005). Selanjutnya model ini disebut dengan model kalibrasi.

Beberapa penulis telah mengembangkan model kalibrasi untuk kasus yang berbeda. Atok (2005) menggunakan Jaringan Syaraf Tiruan dengan metode pra pemrosesan Analisis Komponen Utama, sedangkan Djuraidah (2003) membandingkan kinerja model PLS non-linear dengan Jaringan Syaraf Tiruan pada model kalibrasi. Erfiani (2005) mengembangkan model kalibrasi dengan pendekatan Bayes dimana reduksi peubahnya melalui pendekatan regresi terpenggal, sedangkan Sony (2005) menggunakan Regresi Komponen Utama dimana metode wavelet digunakan untuk pra pemrosesan. Selain pendekatan parametrik, beberapa penulis juga mengembangkan model kalibrasi dengan pendekatan non parametrik. Tonah (2005) menggunakan regresi sinyal P-Spline untuk kalibarasi kandungan gingerol dimana metode prapemrosesannya adalah koreksi pencaran multiplikatif.

Pada penelitian ini penulis menggunakan pendekatan regresi proses Gaussian untuk membangun model kalibrasi pada pengukuran konsentrasi kurkumin berdasarkan data persen transmitannya. Aspek penting yang harus diketahui dalam pemodelan dengan pendekatan regresi proses Gaussian adalah fungsi peragam. Fungsi peragam adalah sebuah fungsi dari input-input model yang menghasilkan sebuah nilai peragam bagi output-output yang bersesuaian (Rasmussen, 1996). Regresi proses Gaussian pada awalnya diusulkan oleh O'Hagan (1978) yang memandang sebagai sebuah alternatif pendekatan untuk jaringan syaraf tiruan. Regresi proses Gaussian dapat juga diturunkan dari perspektif regresi nonparametrik Bayesian dengan penempatan secara langsung sebaran prior Gaussian bagi fungsi-fungsi regresi $f(\mathbf{x})$ (MacKay 1998, diacu dalam Williams 2002). Penelitian ini bertujuan untuk menerapkan regresi proses Gaussian pada pemodelan kalibrasi peubah ganda dengan melakukan kajian terhadap penggunaan berbagai fungsi peragam dan pemilihan bilangan gelombang spektra infra merah dari suatu senyawa aktif.

Regresi Proses Gaussian

Proses stokastik adalah suatu kumpulan dari peubah-peubah acak $\{Y_{\mathbf{x}} | \mathbf{x} \in X\}$ yang diindekskan dengan sebuah himpunan X yang beranggotakan d peubah. Proses-proses stokastik ditentukan oleh pemberian sebaran peluang bersama untuk setiap himpunan bagian manapun dari $Y_{\mathbf{x}_1}, \ldots, Y_{\mathbf{x}_k}$ dengan sebuah cara yang konsisten. Proses Gaussian adalah suatu proses stokastik dimana himpunan berhingga manapun dari himpunan peubah acak Y mempunyai sebaran bersama Gaussian ganda (Williams, 2002). Sebuah proses Gaussian secara lengkap ditentukan oleh fungsi rataan $\mu(\mathbf{x}) = E[Y_{\mathbf{x}}]$ dan fungsi peragam $k(\mathbf{x}_i, \mathbf{x}_j) = E[(Y_{\mathbf{x}_i} - \mu(\mathbf{x}_i))(Y_{\mathbf{x}_j} - \mu(\mathbf{x}_j))]$.

Regresi proses Gaussian dapat diturunkan dari sudut pandang regresi nonparametric Bayesian yaitu dengan penempatan secara langsung sebaran prior Gaussian bagi fungsi-fungsi regresi $f(\mathbf{x})$ (MacKay 1998, diacu dalam Williams 2002). Misal untuk setiap output y_i bergantung pada input $\mathbf{x}_i$ dibawah sebuah fungsi f_i sebagai berikut :

$$y_i = f(\mathbf{x}_i) + \varepsilon_i \tag{1}$$

dimana ε_i adalah peubah acak galat yang secara bebas dan identik menyebar Gaussian dengan rataan nol dan ragam σ^2, sedangkan $\mathbf{x}_i$ adalah vektor input ke-i dimana i = 1,......,n. Apabila fungsi-fungsi f dikumpulkan dalam sebuah vektor $\mathbf{f}$ = $[f_1, \cdots, f_n]^T$ maka menurut Proses Gaussian untuk metode regresi , sebaran prior atas vektor **f** adalah Gaussian Ganda dengan vektor rataan **0** dan matrik peragam **K**, yaitu

$$\mathbf{f}|\mathbf{X},\boldsymbol{\theta} \sim N(\mathbf{0},\mathbf{K}) \qquad (2)$$

dimana **K** adalah matrik n x n yang bergantung pada **X** dan $\boldsymbol{\theta}$ sedangkan $\boldsymbol{\theta}$ adalah vektor parameter dari fungsi peragam. Setiap elemen ke (i,j) dari matrik **K** adalah $k(\mathbf{x}_i,\mathbf{x}_j)$ dimana $k(.,.)$ adalah sebuah fungsi yang definit non negatif yang memuat parameter $\boldsymbol{\theta}$. Selanjutnya $k(.,.)$ disebut sebagai fungsi peragam.

Persamaan (1) dapat dinyatakan dalam bentuk persamaan vektor, yaitu

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\varepsilon} \qquad (3)$$

dimana **y** adalah vektor amatan dari respon, **f** adalah vektor dari fungsi-fungsi regresi dan $\boldsymbol{\varepsilon}$ adalah vektor galat.

Sebagai implikasi langsung atas penetapan sebaran prior Gaussian ganda bagi vektor **f** dan asumsi bahwa vektor galat $\boldsymbol{\varepsilon}$ menyebar Gaussian dengan nilai tengah **0** maka sebaran bagi vektor amatan **y** adalah Gaussian ganda dengan nilai tengah **0** dan matrik ragam peragam $\mathbf{K} + \sigma^2\mathbf{I}$. Tidak setiap vektor amatan **y** selalu memiliki nilai tengah **0** sehingga untuk memenuhinya setiap amatan dari y_i akan dikurangi dengan nilai rata-rata dari keseluruhan amatan.

Fungsi Peragam

Fungsi peragam adalah sebuah fungsi dari input-input model yang menghasilkan sebuah nilai peragam bagi output-output yang bersesuaian (Rasmussen, 1996). Satu-satunya syarat bagi sebuah fungsi peragam adalah mampu membangkitkan sebuah matrik ragam peragam yang definit non negatif untuk sembarang himpunan titik-titik input.

Beberapa fungsi peragam yang umum digunakan dalam model regresi proses Gaussian adalah sebagai berikut:

a. Fungsi peragam kuadrat eksponensial dengan ukuran jarak isotropik (KE-iso).
b. Fungsi peragam kuadrat eksponensial dengan ukuran jarak Automatic Relevance Determination (KE-ard).
c. Fungsi peragam linear dengan hiperparameter tunggal (Linear-1).

d. Fungsi peragam linear dengan parameter Automatic Relevance Determination (Linear-ard).
e. Fungsi peragam Matern 3 dengan ukuran jarak isotropik.

Fungsi peragam lainnya dapat dibentuk dengan mengkombinasikan fungsi-fungsi peragam di atas (Rasmussen dan Williams, 2006). Dalam setiap fungsi peragam mengandung beberapa parameter yang selanjutnya disebut dengan hiperparameter.

Pendugaan Hiperparameter Fungsi Peragam

Terdapat beberapa metode yang dapat digunakan untuk menduga nilai-nilai hiperparameter. Williams (2002) menyatakan bahwa untuk menduga nilai $\boldsymbol{\theta}$ dapat digunakan metode kemungkinan marginal maksimum (*Maximum Marginal Likelihood*) , metode aposterior maksimum, dan metode simulasi hybrid Monte Carlo. Metode lain yang bisa digunakan adalah metode *Cross Validation* dan metode *Generalized Cross Validation* (Wahba, 1990 dalam Williams 2002).

Dalam penelitian ini, pendugaan nilai hiperparameter menggunakan metode kemungkinan marginal maksimum. Fungsi kemungkinan marginal diperoleh dengan mengintegralkan fungsi kemungkinan yang telah dikalikan dengan sebaran prior bagi **f**, yaitu

$$p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta}) = \int p(\mathbf{y}|\mathbf{f},\mathbf{X},\boldsymbol{\theta})p(\mathbf{f}|\mathbf{X},\boldsymbol{\theta})d\mathbf{f} \qquad (4)$$

Dibawah kerangka kerja Proses Gaussian sebaran prior atas $\mathbf{f}|\mathbf{X},\boldsymbol{\theta}$ adalah Gaussian ganda, yaitu $\mathbf{f}|\mathbf{X},\boldsymbol{\theta} \sim N(\mathbf{0},\mathbf{K})$ atau

$$\log p(\mathbf{f}|\mathbf{X},\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f} - \frac{1}{2}\log|\mathbf{K}| - \frac{n}{2}\log 2\pi \qquad (5)$$

Peubah acak $\mathbf{y}|\mathbf{f},\mathbf{X},\boldsymbol{\theta}$ dan $\mathbf{f}|\mathbf{X},\boldsymbol{\theta}$ masing-masing menyebar Gaussian Ganda sehingga peubah acak $\mathbf{y}|\mathbf{X},\boldsymbol{\theta}$ menyebar Gaussian ganda juga (Timm, 2002), sehingga fungsi kemungkinan marginalnya menurut Rasmussen (2006) adalah

$$\log p(\mathbf{y}|\mathbf{X},\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T\left(\mathbf{K}+\sigma^2\mathbf{I}\right)^{-1}\mathbf{y} - \frac{1}{2}\log\left|\mathbf{K}+\sigma^2\mathbf{I}\right| - \frac{n}{2}\log 2\pi \qquad (6)$$

Penduga bagi hiperparameter fungsi peragam tidak dapat diperoleh secara langsung melalui statistik penduganya, oleh karena itu untuk menemukan nilai dugaannya dilakukan secara numerik. Salah satu metode yang dapat digunakan adalah metode *Conjugate Gradient* (Fletcher dan Reeves, 1964). Metode *Conjugate gradient* adalah sebuah algoritma yang dirancang untuk menemukan nilai minimum lokal terdekat dari fungsi banyak peubah dengan syarat gradien dari fungsi tersebut dapat dihitung. Usaha untuk memaksimumkan fungsi kemungkinan marginal ekuivalen dengan meminimumkan fungsi kemungkinan marginal negatif.

Prediksi Dalam Regresi Proses Gaussian

Misal diberikan beberapa amatan dan sebuah fungsi peragam. Selanjutnya akan ditentukan nilai prediksinya dengan menggunakan model proses Gaussian. Untuk melakukan hal itu, jika $\mathbf{x}^*$ sebuah titik uji dan f^* adalah fungsi yang bersesuaian dengan $\mathbf{x}^*$, maka dibawah kerangka kerja Proses Gaussian , sebaran bersama dari **f** dan f^* adalah Gaussian Ganda dengan rataan nol, yaitu:

$$\begin{bmatrix} \mathbf{f} \\ f^* \end{bmatrix} \Bigg| \mathbf{X}, \boldsymbol{\theta} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{k} \\ \mathbf{k}^{\mathbf{T}} & \kappa \end{bmatrix}\right) \tag{7}$$

dimana $\mathbf{k} = [k(\mathbf{x}^*, \mathbf{x}_1), \ldots, k(\mathbf{x}^*, \mathbf{x}_n)]^T$ adalah vektor n x 1 yang dibentuk dari peragam antara $\mathbf{x}^*$ dan input-input model. Sedangkan $\kappa = k(\mathbf{x}^*, \mathbf{x}^*)$adalah sebuah skalar. Apabila peubah galat mengikuti sebaran seperti pada persamaan (1) maka sebaran bersama dari nilai teramati **y** dan y^* adalah

$$\begin{bmatrix} \mathbf{y} \\ y^* \end{bmatrix} \Bigg| \mathbf{X}, \boldsymbol{\theta} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma^2 I & \mathbf{k} \\ \mathbf{k}^{\mathbf{T}} & \kappa + \sigma^2 \end{bmatrix}\right) \tag{8}$$

Sehingga sebaran marginal dari y^* adalah Gaussian juga, yaitu :

$$y^* | \mathbf{y}, \mathbf{X}, \boldsymbol{\theta} \sim N(m(\mathbf{x}^*), v(\mathbf{x}^*)) \tag{9}$$

dimana rataan dan ragam adalah

$$m(\mathbf{x}^*) = \mathbf{k}^{\mathbf{T}} (\mathbf{K} + \sigma^2 \mathbf{I})^{-1} \mathbf{y} \tag{10}$$

$$v\left(\mathbf{x}^*\right) = \kappa + \sigma^2 - \mathbf{k}^{\mathbf{T}}\left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{k} \tag{11}$$

Nilai dugaan bagi y^* adalah $m(\mathbf{x}^*)$ dan ragam bagi dugaan y^* adalah $v\left(\mathbf{x}^*\right)$. Secara umum untuk m buah titik uji $\mathbf{X}^* = \left[\mathbf{x}_1^*, \ldots, \mathbf{x}_m^*\right]$ maka sebaran $\mathbf{y}^*$ adalah Gaussian Ganda dengan parameter-parameter,

$$m\left(\mathbf{X}^*\right) = \mathbf{K}^{*\mathbf{T}}\left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{y} \tag{12}$$

$$v\left(\mathbf{X}^*\right) = \mathbf{K}^{**} + \sigma^2\mathbf{I} - \mathbf{K}^{*\mathbf{T}}\left(\mathbf{K} + \sigma^2\mathbf{I}\right)^{-1}\mathbf{K} \tag{13}$$

dimana $\mathbf{K}^*$ adalah matrik n x m dari peragam antara input-input training dan titik-titik uji. Matrik $\mathbf{K}^{**}$ dengan ukuran m x m tersusun dari peragam antara titik-titik uji.

Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang merupakan bagian dari data penelitian Hibah Pascasarjana tahun 2003-2005 hasil kerjasama antara Departemen Statistika IPB dengan Pusat Studi Biofarmaka LPPM IPB. Penelitian tersebut didanai oleh Dirjen Pendidikan Tinggi, Departemen Pendidikan Nasional. Data yang digunakan adalah persen transmitan kurkumin dari serbuk temulawak hasil pengukuran spektrometer FTIR dan data konsentrasi senyawa aktif kurkumin yang diukur dengan menggunakan HPLC. Temulawak yang dijadikan contoh diambil dari beberapa daerah sentra tanaman obat, yaitu Bogor, Sukabumi, Kulon Progo, Karanganyar, dan Cianjur dan Balitro. Data-data tersebut diperoleh dari Pusat Studi Biofarmaka Institut pertanian Bogor.

Hasil dan Pembahasan

Data persen transmitan dalam penelitian ini berperan sebagai peubah penjelas yang nilai-nilainya diukur pada daerah identifikasi gugus fungsional kurkuminoid dimana bilangan-bilangan gelombangnya dapat dilihat di Socrates (2004). Daerah identifikasi adalah daerah yang menunjukkan absorpsi setiap gugus fungisonal yang disebabkan oleh adanya vibrasi regangan. Karena peubah penjelas yang digunakan jumlahnya lebih banyak jika dibandingkan dengan jumlah contoh yang digunakan maka jumlahnya perlu direduksi. Analisis Komponen Utama (AKU) digunakan untuk mereduksi banyaknya peubah penjelas dengan persentase keragaman kumulatif yang mampu dijelaskan digunakan sebagai kriteria untuk menentukan banyaknya komponen utama. Tabel 1 menjelaskan bahwa dengan menggunakan 1 komponen utama, keragaman yang dapat dijelaskan sebesar 90,20% dan apabila menggunakan 2 komponen utama keragaman yang dapat dijelaskan sebesar 97,22% sedangkan apabila menggunakan 3 komponen utama keragaman yang dapat dijelaskan sebesar 99,66% dari keragaman pada data asal. Oleh karena itu dalam analisis selanjutnya digunakan 3 komponen utama pertama sebagai peubah penjelas.

Tabel 1 Ragam kumulatif komponen utama

Komponen	Ragam	Prosentase	Ragam Kumulatif
1	11.9088	90.20%	90.20%
2	0.9263	7.02%	97.22%
3	0.3228	2.44%	99.66%
4	0.0289	0.22%	99.88%
5	0.0087	0.07%	99.95%

Fungsi-fungsi peragam yang dapat dipilih dalam pemodelan regresi proses Gaussian dapat dilihat dalam tulisan Rassmussen dan Williams (2006).

Tabel 2 Nilai RMSEP setiap jenis fungsi peragam

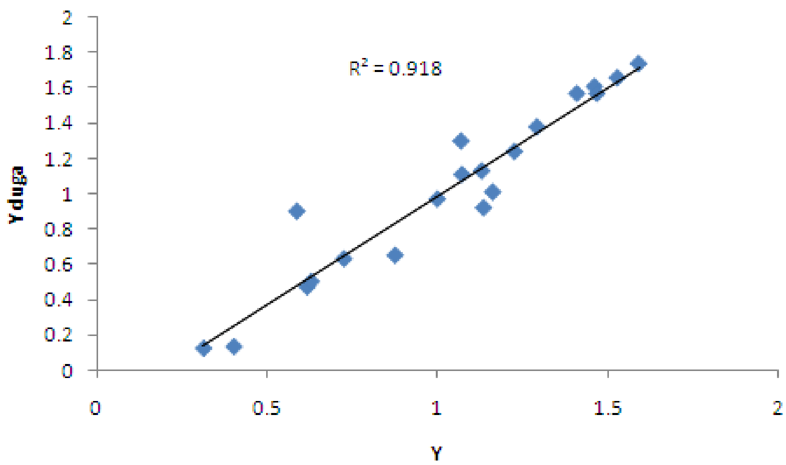
No	Fungsi Peragam	Rata-Rata RMSEP	
		Bil. Gelombang 4000-400 cm^{-1}	Daerah Identifikasi
1	Kuadrat Eksponensial - Isotropik (KE-iso)	0,5913	0,5242
2	Kuadrat Eksponensial - Automatic Relevance Determinant (KE - ard)	0,5446	0,5382
3	Linear 1	0,6963	0,7033
4	Linear Automatic Relevance Determinant (Linear - ard)	0,6540	0,6701
5	Matern 3 - isotropik	0,5234	0,5179
6	Linear 1 + KE - iso	0,6350	0,6423
7	Linear ard + KE - iso	0,6444	0,5658
8	Linear 1 + KE - ard	0,6444	0,6587
9	Linear ard + KE - ard	0,6268	0,6240

Dengan menggunakan skor dari 3 komponen utama yang telah ditetapkan pada bagian sebelumnya dan menganggap konsentrasi kurkumin hasil dari pengukuran HPLC sebagai peubah respon maka pemilihan fungsi peragam dilakukan dengan memperhatikan nilai RMSEP yang dihasilkanya. Fungsi peragam yang menghasilkan nilai RMSEP terkecil akan dipilih menjadi fungsi peragam bagi model kalibrasi konsetrasi kurkumin.

Dari Tabel 2 dapat disimpulkan bahwa fungsi peragam yang relevan untuk pemodelan kalibrasi konsentrasi kurkumin pada daerah identifikasi gugus fungsional kurkuminoid adalah Matern 3 isotropik karena memberikan nilai RMSEP terkecil, yaitu sebesar 0,5178. Dengan menggunakan fungsi peragam yang sama, pemodelan kalibrasi untuk pengukuran konsentrasi kurkumin pada pemilihan bilangan gelombang antara 4000-400 cm^{-1} menghasilkan nilai RMSEP yang lebih besar , yaitu 0,5234. Fungsi peragam matern 3 isotropik memiliki formula matematis sebagai berikut:

$$k(\mathbf{x}_i,\mathbf{x}_j)=\sigma_f^2\left(1+\sqrt{3(\mathbf{x}_i-\mathbf{x}_j)^T\Delta^{-1}(\mathbf{x}_i-\mathbf{x}_j)}\right)\exp\left(-\sqrt{3(\mathbf{x}_i-\mathbf{x}_j)^T\Delta^{-1}(\mathbf{x}_i-\mathbf{x}_j)}\right)$$

, dimana σ_f^2 adalah ragam signal dan $\Delta=\begin{bmatrix} l_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & l_d^2 \end{bmatrix}$ sedangkan l_m adalah parameter skala panjang untuk m = 1, ..., d. Dalam fungsi peragam ini nilai parameter skala panjang dianggap sama yaitu $l_1=l_2=\cdots=l_d=l$.

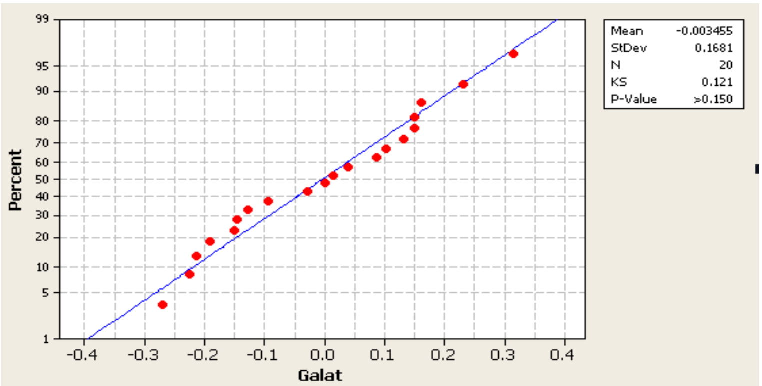


Gambar 1 Plot antara Y dan Y duga untuk model regresi proses Gaussian

Dengan menggunakan metode kemungkinan marginal maksimum diperoleh nilai dugaan bagi σ_f^2 dan l yang masing-masing sebesar 0,2244 dan 1,0606. Implementasi regresi proses Gaussian dengan fungsi peragam Matern 3 isotropik menghasilkan nilai $R^2_{Y vs \hat{Y}}$ sebesar 91,8% dengan nilai RMSE sebesar 0,1639. Gambar 1 adalah plot antara nilai aktual konsentrasi kurkumin dan nilai dugaannya dibawah model regresi proses Gaussian. Tampak bahwa plot diantara keduanya cenderung membentuk garis lurus, meskipun garisnya tidak melalui pusat koordinat.

Munculnya asumsi sebaran Gaussian bagi peubah acak galat pada regresi proses Gaussian memiliki tujuan yang berbeda dengan munculnya asumsi sebaran Gaussian bagi peubah acak galat pada regresi parameterik pada umumnya. Pada regresi parametrik adanya asumsi tersebut berguna untuk pengujian hipotesis bagi parameter-perameter model regresinya sedangkan pada regresi proses Gaussian adanya asumsi tersebut semata-mata agar sebaran bagi amatan *y* dapat ditelusuri.

Gambar 2 menunjukkan plot peluang normal dari peubah acak galat. Tampak bahwa sebagian besar data menyebar disepanjang garis lurus. Hal ini mengindikasikan bahwa data menyebar normal. Dengan menggunakan uji Kolmogorov-Smirnov, pada pemilihan $\alpha = 5\%$ diketahui bahwa peubah acak galat mengikuti sebaran normal (p-value > 0,150). Oleh karena itu asumsi yang dibutuhkan dalam pemodelan regresi proses Gaussian ini telah terpenuhi.



Gambar 2 Plot peluang normal peubah acak galat

Kesimpulan

Berdasarkan nilai RMSEP, fungsi peragam yang relevan untuk pemodelan kalibrasi pengukuran konsentrasi kurkumin pada daerah identifikasi spektra infra merah dengan pendekatan regresi proses Gaussian adalah Matern 3 isotropik. Nilai RMSEP yang diperoleh sebesar 0,5179. Dengan menggunakan fungsi peragam yang sama, nilai RMSEP yang dicapai oleh model kalibrasi pengukuran konsentrasi kurkumin pada bilangan gelombang 4000-400 cm^{-1} sebesar 0,5234. Hal itu berarti lebih besar daripada nilai yang dicapai jika bilangan gelombang yang dipilih pada daerah identifikasi spektra infra merah gugus fungsional kurkuminoid. Namun tidak semua fungsi peragam mempunyai kinerja yang lebih baik pada daerah tersebut.

DAFTAR PUSTAKA

Atok RM. 2005. Jaringan Syaraf Tiruan dalam Pemodelan Kalibrasi dengan Pra-pemrosesan Analisis Komponen Utama dan Transformasi Fourier Diskrit [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.

Chen T, Morris J, Martin E. 2007. Gaussian Process Regression for Multivariate Spectroscopic Callibration. *Chemometrics and Intelligent Laboratory Systems* 87: 85-97.

Djuraidah A. 2003. Penerapan Model Nonlinear PLS dengan Jaringan Syaraf Tiruan dalam Kalibrasi. *Jurnal Matematika Aplikasi dan Pembelajarannya* (*JMAP*) 2:339-345.

Erfiani. 2005. Pengembangan Model Kalibrasi dengan Pendekatan Bayes (Kasus Tanaman Obat [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.

Fletcher R, Reeves CM.1964. Function Minimization by Conjugate Gradients. *Computer Journal* 7:148–154.

Naes T, Issackson T, Fearn T, Davies T. 2002. *User Friendly Guide to Multivariate Calibration and Classification*. United Kingdom: NIR Publication Chichester.

Neal RM. 1996.*Bayesian learning for neural network*. New York:Springer-Verlag.

Nur MA, Adijuwana H. 1989. *Teknik Spektroskopi dalam Analisis Biologi*. Bogor: Pusat Antar Univrsitas Ilmu Hayat, Institut Pertanian Bogor.

O'Hagan. 1978. Curve fitting and optimal design for prediction (with discussion). *Journal of the Royal Statistical Society B 40*. 1-40.

Rasmussen CE. 1996. Evaluation of Gaussian Processes and other Methods for Non-linear Regression [Disertasi]. Toronto: Department of Computer Science, University of Toronto.

Rasmussen CE, Williams CKI. 2006. *Gaussian Process for Machine Learning*. Massachusetts : MIT Press.

Socrates G. 1994. *Infared Characteristic Group Frequencies Tables and Charts*. Edisi Ke-2. England : John Wiley and Sons.

Sunaryo S. 2005. Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-pemrosesan [Disertasi]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.

Tonah. 2006. Pemodelan Kalibrasi Peubah Ganda dengan Pendekatan Regresi Sinyal P-Spline [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.

Timm NH. 2002. *Applied Multivariate Analysis*. New York. Springer.