

PEMODELAN KALIBRASI PEUBAH GANDA DENGAN PENDEKATAN REGRESI SINYAL P-SPLINE (*Multivariate Calibration Models using P-spline Signal Regression Approach*)

Tonah¹⁾, Ahmad Ansori Mattjik²⁾, Khairil Anwar Notodiputro²⁾

¹⁾ Prodi Pendidikan Matematika FKIP

Universitas Swadaya Gunung Jati Cirebon

²⁾ Departemen Statistika FMIPA Institut Pertanian Bogor

E-mail : tonah_spsstk02@yahoo.com

Abstrak

Model kalibrasi peubah ganda merupakan suatu fungsi hubungan antara satuan pengukuran yang dapat diperoleh melalui proses yang relatif mudah atau murah dengan satuan pengukuran yang memerlukan waktu lama dan biaya mahal. Secara umum data kalibrasi memiliki multikolinearitas yang tinggi antar peubah penjelas dan dimensinya jauh lebih besar daripada banyaknya contoh yang dimiliki. Oleh karena itu, sebagian besar pendekatan model kalibrasi memerlukan pereduksian data terlebih dulu. Solusi alternatif bagi pemodelan kalibrasi adalah regresi sinyal P-spline (RSP). RSP merupakan salah satu pendekatan nonparametrik yang mensyaratkan bahwa koefisien regresi berada dalam ruang fungsi mulus. Hal ini dilakukan dengan cara merepresentasikan koefisien regresi sebagai kombinasi linear dari basis B-spline. Penambahan penalti dilakukan untuk mengatasi multikolinearitas pada model serta meningkatkan kemulusan koefisien regresi. Indeks dari bilangan gelombang yang terukur oleh FTIR digunakan sebagai domain B-spline. Spektra gingerol diidentifikasi memiliki pengaruh pencaran multiplikatif, sehingga perlu dilakukan koreksi pencaran. Model RSP dengan koreksi pencaran multiplikatif pada senyawa aktif gingerol memberikan hasil prediksi yang lebih baik. Hal ini ditunjukkan oleh nilai RMSEP dan $R^2_{y \text{ vs } \hat{y}}$ masing-masing sebesar 0.06867 dan 95.73 %. Nilai-nilai tersebut jauh lebih kecil dari hasil yang diberikan oleh model regresi komponen utama dengan pra-pemrosesan koreksi pencaran maupun transformasi wavelet.

Kata kunci : model kalibrasi, koreksi pencaran, B-spline, regresi sinyal P-spline

PENDAHULUAN

Model kalibrasi peubah ganda merupakan suatu fungsi hubungan antara satuan pengukuran yang dapat diperoleh melalui proses yang relatif mudah atau murah dan satuan pengukuran yang memerlukan waktu lama dan biaya mahal dalam memperolehnya (Naes *et al.* 2002). Ukuran yang mahal tersebut adalah konsentrasi suatu unsur atau senyawa yang dihasilkan oleh HPLC (*High Performance Liquid Chromatography*), sedangkan ukuran-ukuran yang murah adalah persen transmittan pada bilangan gelombang yang dihasilkan oleh spektrometer FTIR (*Fourier Transform Infrared*). Tujuan pemodelan kalibrasi peubah ganda adalah menemukan model yang dapat digunakan untuk memprediksi konsentrasi senyawa

secara akurat berdasarkan informasi persen transmittan dari senyawa yang dianalisis.

Secara umum, pada data kalibrasi terdapat beberapa permasalahan di antaranya yaitu adanya multikolinearitas yang tinggi antar peubah penjelas (pada data persen transmittan) dan banyaknya peubah penjelas jauh lebih besar daripada banyaknya contoh yang dimiliki. Beberapa metode yang digunakan dalam model kalibrasi di antaranya adalah regresi *ridge*, regresi bertatar, pendekatan Bayes (Rahayu 2003, Erfiani 2005), regresi kuadrat terkecil parsial (RKTP) dan regresi komponen utama (RKU), jaringan syaraf tiruan (JST) (Atok 2005, Djuraidah 2003). Dimensi yang dimiliki data kalibrasi sangat besar, sehingga sebagian besar analisis memerlukan pereduksian dimensi data terlebih dulu. Teknik reduksi data yang digunakan dalam model kalibrasi di antaranya adalah

komponen utama, transformasi wavelet diskret (Sunaryo 2005), transformasi fourier diskret (Atok 2005), pendekatan regresi terpenggal (Erfiani 2005), basis B-spline (Alsberg 1993, diacu dalam Marx & Eilers 1999).

Solusi alternatif lain yang dapat digunakan dalam pemodelan kalibrasi adalah regresi sinyal P-spline (RSP). RSP merupakan salah satu pendekatan nonparametrik yang mensyaratkan bahwa koefisien regresi berada dalam ruang fungsi mulus. Hal ini dilakukan dengan cara memroyeksikan koefisien regresi ke dalam basis yang berada dalam ruang fungsi mulus. Basis yang digunakan dalam RSP adalah basis B-spline. RSP merupakan suatu model regresi linear berganda terpenalti yang terkendala pada suatu ruang fungsi mulus. Penambahan penalti bertujuan untuk mengatasi multikolinearitas pada model serta meningkatkan kemulusan koefisien regresi.

Pra-pemrosesan merupakan langkah yang penting dalam model kalibrasi. Pra-pemrosesan dapat meningkatkan kemampuan prediksi dari suatu model, terutama untuk spektrum dengan keragaman pencaran antar spektra yang didominasi oleh pengaruh aditif dan multiplikatif. Arnita (2005) telah mengidentifikasi adanya pengaruh pencaran aditif dan multiplikatif pada data persen transmittan senyawa aktif gingerol serbuk rimpang jahe, kemudian melakukan koreksi pencaran menggunakan koreksi pencaran multiplikatif. Oleh karena itu, kemampuan prediksi dari model yang dibentuk menggunakan RKU dapat ditingkatkan. Berdasarkan hal tersebut, penelitian ini akan menggabungkan pra-pemrosesan koreksi pencaran multiplikatif dengan model RSP pada data yang sama.

Tujuan penelitian ini adalah membandingkan kemampuan prediksi RSP pada data yang terlebih dulu dilakukan pra-pemrosesan koreksi pencaran dengan data tanpa pra-pemrosesan.

KOREKSI PENCARAN MULTIPLIKATIF

Pra-pemrosesan merupakan tahapan yang penting dalam pemodelan kalibrasi, dan salah satu teknik yang sering digunakan adalah koreksi pencaran multiplikatif. Koreksi ini dapat mengeliminasi keragaman antar spektra yang didominasi oleh pengaruh aditif dan multiplikatif sehingga dapat meningkatkan kelinearan dan kemampuan prediksi dari suatu model (Naes *et al* 2002). Koreksi pencaran multiplikatif dilakukan dengan meregresikan spektrum masing-masing contoh dengan spektrum rataannya. Bentuk persamaan regresi untuk masing-masing contoh sebagai berikut

$$x_{ij} = \beta_{0i} + \beta_{1i}\bar{x}_j + e_i \quad i=1,2,\dots,n; \quad j=1,2,\dots,p \quad (1)$$

dimana x_{ij} adalah spektrum contoh ke- i pada bilangan gelombang ke- j , β_{0i} adalah pengaruh

aditif (intersep) pada contoh ke- i , β_{1i} adalah pengaruh multiplikatif (kemiringan) pada contoh ke- i , \bar{x}_j adalah rata-rata seluruh spektrum pada bilangan gelombang ke- j .

Pengaruh aditif dan multiplikatif pada setiap contoh diduga dengan menggunakan metode kuadrat terkecil. Setelah diperoleh nilai dugaannya, kemudian digunakan untuk mentransformasi spektrum asli dengan menggunakan persamaan berikut

$$x_{ij}^* = (x_{ij} - \beta_{0i}) / \beta_{1i} \quad (2)$$

Spektrum yang sudah dikoreksi dari masing-masing contoh memiliki posisi pencaran yang sama. Hal ini mengindikasikan bahwa perbedaan informasi kimia yang diberikan antar contoh relatif kecil (De Maesschalck *et al* 1999 dan Naes *et al* 2002, diacu dalam Arnita 2005).

FUNGSI BASIS B-SPLINE

Fungsi spline merupakan potongan polinomial yang memiliki ruas-ruas polinomial yang berbeda dan tersambung bersama pada titik-titik *knot* dengan syarat jaminan kekontinuan tertentu (Gunawan 2001). Setiap fungsi spline berderajat q dengan barisan *knot* T dapat dituliskan sebagai kombinasi linear basis B-spline (De Boor 1978). Misalkan $T = \{t_0 \leq t_1 \leq \dots \leq t_m\}$ adalah himpunan $(m+1)$ *knot* pada selang $[t_0, t_m]$, fungsi basis B-spline berderajat q ke- i ($B_{i,q}$) didefinisikan secara rekursif (rekursif Carl De Boor) dengan persamaan berikut

$$B_{i,0}(x) = \begin{cases} 1, & \text{jika } t_i \leq x \leq t_{i+1} \\ 0, & \text{selainnya} \end{cases} \quad \text{dan}$$

$$B_{i,q}(x) = \frac{x-t_i}{t_{i+q}-t_i} B_{i,q-1}(x) + \frac{t_{i+q+1}-x}{t_{i+q+1}-t_{i+1}} B_{i+1,q-1}(x).$$

REGRESI SINYAL P=SPLINE

RSP merupakan salah satu pendekatan nonparametrik yang mensyaratkan bahwa koefisien regresi berada dalam ruang fungsi mulus. Hal ini dilakukan dengan cara memroyeksikan koefisien regresi ke dalam basis fungsi mulus. Koefisien regresi dinyatakan sebagai hasil kali suatu vektor koefisien basis dengan matriks basis. Basis yang digunakan dalam RSP adalah basis B-spline. Indeks bilangan gelombang merupakan bagian yang sangat penting dalam RSP sebagai dasar (domain) dalam membangun basis B-spline.

Model kalibrasi dapat dipandang sebagai regresi linear yang dituliskan dalam bentuk

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \dots + \alpha_p x_{pi} + e_i \quad , \quad (i=1,2,\dots,n).$$

Nilai harapan persamaan di atas adalah

$$E(\mathbf{y}_n) = \alpha_0 \mathbf{1}_n + \mathbf{X}_{(n \times p)} \boldsymbol{\alpha}_p \quad (3)$$

dimana \mathbf{X} adalah matriks spektrum (peubah penjelas) yang dihasilkan pada p bilangan gelombang, antar peubah penjelas terdapat multikolinearitas, $p \gg n$. Matriks \mathbf{X} memiliki dimensi yang sangat besar sehingga dimensi vektor koefisien parameter juga besar. Oleh karena itu, perlu dilakukan pereduksian dimensi pada vektor koefisien regresi $\boldsymbol{\alpha}$ dengan cara memroyeksikan $\boldsymbol{\alpha}$ pada basis fungsi mulus berdimensi sedang yaitu basis B-spline (\mathbf{B}). $\boldsymbol{\alpha}$ dapat diformulasikan sebagai

$$\boldsymbol{\alpha}_p = \mathbf{B}_{(p \times m)} \boldsymbol{\beta}_m \quad (4)$$

dengan $\boldsymbol{\beta}$ adalah vektor parameter koefisien basis B-spline berdimensi m .

Permasalahan mendasar dalam membangun fungsi basis B-spline adalah penentuan jumlah dan penempatan *knot* yaitu tempat tersambungannya potongan-potongan polinomial pada B-spline. Jumlah *knot* yang terlalu banyak akan menyebabkan *overfitting*, sebaliknya jika terlalu sedikit akan mengakibatkan *underfitting*. Pengoptimalan jumlah dan penempatan *knot* merupakan masalah non linear yang kompleks dan lebih mengarah pada masalah algoritma sehingga tidak dijadikan pembahasan dalam tulisan ini.

P-spline dapat mengatasi permasalahan di atas dengan menentukan terlebih dulu jumlah *knot* yang akan digunakan pada B-spline, dan penempatan *knot* dilakukan dengan konsep *equally spaced knots* yaitu mengatur posisi *knot* sedemikian rupa sehingga antar *knot* memiliki jarak yang sama. Kemudian untuk meningkatkan kemulusan diberikan penalti berbeda pada koefisien-koefisien $\boldsymbol{\beta}$ yang berdekatan.

Dengan mensubstitusikan persamaan (4) ke persamaan (3) diperoleh

$$\mathbf{E}(\mathbf{y}_n) = \alpha_0 \mathbf{1}_n + \mathbf{U}_{(n \times m)} \boldsymbol{\beta}_m \quad (5)$$

dengan matriks $\mathbf{U}_{(n \times m)} = \mathbf{X}_{(n \times p)} \mathbf{B}_{(p \times m)}$, dengan $m \ll p$. Selanjutnya peningkatan pemulusan dilakukan dengan menambahkan penalti pembeda (λ) dan penalti *ridge* (κ) pada $\boldsymbol{\beta}$. Penambahan penalti tersebut sekaligus merupakan solusi bagi permasalahan multikolinearitas pada data. Model yang akan dibangun diharapkan memiliki *doubly penalized least square* (S) yang minimum.

$$S = |\mathbf{y} - \alpha_0 - \mathbf{U}\boldsymbol{\beta}|^2 + \lambda \sum_{k=d+1}^m (\Delta_k^d \boldsymbol{\beta})^2 + \kappa \sum_{i=1}^m \beta_i^2 \quad (6)$$

bentuk matriks dari persamaan di atas adalah:

$$S = \mathbf{y}^T \mathbf{y} - 2\boldsymbol{\beta}^* \mathbf{U}^{*T} \mathbf{y} + \boldsymbol{\beta}^{*T} \mathbf{U}^* \mathbf{U}^* \boldsymbol{\beta}^* + \lambda \boldsymbol{\beta}^{*T} \mathbf{D}_d^{*T} \mathbf{D}_d^* \boldsymbol{\beta}^* + \kappa \boldsymbol{\beta}^{*T} \mathbf{I}_{(m+1)}^* \boldsymbol{\beta}^*$$

dengan $\mathbf{U}^* = (\mathbf{1}_n | \mathbf{U})$, $\mathbf{D}_d^* = (\mathbf{0}_{(m-d)} | \mathbf{D}_d)$, $\mathbf{I}^* = \text{diag}(0, \mathbf{1}_m)$

dan $\boldsymbol{\beta}^{*T} = (\alpha_0 | \boldsymbol{\beta}^T)$. \mathbf{D}_d adalah matriks tanda berukuran $(m-d) \times m$ dihitung secara rekursif, dengan persamaan berikut

$$\mathbf{D}_0 \boldsymbol{\beta} = \boldsymbol{\beta} ; \mathbf{D}_1 \boldsymbol{\beta} = \{\beta_k - \beta_{k-1}\};$$

$$\mathbf{D}_{d+1} \boldsymbol{\beta} = \mathbf{D}_1 \mathbf{D}_d \boldsymbol{\beta} \text{ untuk } k = 2, 3, \dots, m.$$

dan $\lambda \geq 0$, $\kappa \geq 0$ adalah konstanta yang sangat kecil).

Pendugaan parameter $\boldsymbol{\beta}^*$, λ dan κ yang meminimumkan S dilakukan secara serentak menggunakan algoritma Newton-Rapshon. Nilai prediksi (\hat{y}) konsentrasi gingerol kelompok data validasi dapat diperoleh dari persamaan berikut

$$\begin{aligned} \hat{\mathbf{y}} &= \mathbf{X} \hat{\boldsymbol{\alpha}}^* \\ \hat{\boldsymbol{\alpha}}^* &= \mathbf{B} \hat{\boldsymbol{\beta}}^* \end{aligned} \quad (7)$$

Ukuran kebaikan hasil prediksi yang digunakan adalah nilai R^2 dari regresi \hat{y} terhadap y dan RMSEP (*Root Mean Square Error of Prediction*)

$$\text{RMSEP} = \left\{ \frac{1}{n_c} \sum_{i=1}^{n_c} (y_i - \hat{y}_i)^2 \right\}^{1/2} \quad (8)$$

dimana n_c adalah banyaknya kelompok contoh validasi.

METODOLOGI PENELITIAN

Sumber Data

Data yang akan digunakan dalam penelitian ini berasal dari data penelitian Hibah Pascasarjana Pengembangan Model untuk Pendugaan Senyawa Bioaktif atau Senyawa Penciri Beberapa Tanaman Obat, SPS-IPB. Data ini berupa data persen transmitan hasil pengukuran spektrometer FTIR dan data konsentrasi hasil pengukuran HPLC pada senyawa aktif *gingerol* serbuk rimpang jahe hasil survey pada para petani di Kulon Progo, Karanganyar, Majalengka, Bogor dan jahe hasil koleksi dari Balitro.

Metode

Tahapan analisis data yang dilakukan pada penelitian ini adalah sebagai berikut :

- A. Tahap pra-pemrosesan meliputi
 1. Spektrum dari semua contoh di plot terhadap rata-rata untuk melihat pola garis lurus yang dibentuk oleh tiap contoh tersebut.
 2. Membentuk model regresi linear $x_{ij} = \beta_{0i} + \beta_{1i} \bar{x}_j + e_i$ dan menduga nilai β_{0i} dan β_{1i} menggunakan metode kuadrat terkecil.
 3. Mentransformasi data spektrum asli dengan menggunakan transformasi berikut: $x_{ij}^* = (x_{ij} - \hat{\beta}_{0i}) / \hat{\beta}_{1i}$.
 4. Data hasil transformasi dibagi secara acak ke dalam dua kelompok yaitu data untuk membangun model (data kalibrasi) dan data untuk validasi model. Pada penelitian ini data yang digunakan terdiri dari dua puluh contoh. Lima belas contoh digunakan untuk membangun model dan lima contoh sisanya digunakan untuk validasi model.

B. Tahap pembentukan dan validasi model mencakup

1. Ditetapkan sejumlah *knot* dengan jarak yang sama. Misalkan domain B-spline kubik terbagi pada m' interval yang sama panjang yang terbagi oleh $m'+1$ *knot* internal ($h = t_{i+1} - t_i$, untuk $3 \leq i \leq m'+2$ dan $h \in \mathfrak{R}$).

Jumlah total *knot* yang dibutuhkan untuk membangun B-spline = $m'+7$ dan banyaknya B-spline yang digunakan pada persamaan regresi $m = m'+3$. Pada Penelitian ini B-spline kubik yang dibentuk akan dicobakan pada 12 interval dan 13 interval *knot*.

2. Membangun fungsi basis B-spline kubik sepanjang indeks bilangan gelombang (1:p) dengan sejumlah *knot* yang diberikan pada Langkah 1. Konsentrasi senyawa aktif gingerol pada serbuk rimpang jahe dipengaruhi oleh masa penyimpanannya, sehingga dalam model yang akan dibangun ditambahkan peubah *dummy* masa penyimpanan (I).

$$I = \begin{cases} 1, & \text{masa simpan sebentar (kurang dari tiga bulan)} \\ 0, & \text{masa simpan lama (lebih dari tiga bulan)} \end{cases}$$

Peubah *dummy* ini tidak dilibatkan dalam pemulusan koefisien regresi, sehingga domain dari B-spline kubik yang akan dibentuk hanya indeks bilangan gelombang yaitu $\{1, 2, \dots, 1866\}$. Peubah *dummy* akan diikuti dalam pemodelan dengan cara memperbesar matriks-matriks berikut

$$\mathbf{X}_{(n \times (p+1))} = [\mathbf{X}_{(n \times p)} : \mathbf{I}_n];$$

$$\mathbf{B}_{((p+1) \times (m+1))} = \begin{bmatrix} \mathbf{B}_{(p \times m)} & \mathbf{0}_p \\ \mathbf{0}_m^T & 1 \end{bmatrix};$$

$$\mathbf{I}_{((m+1) \times (m+1))} = \text{diag}(\mathbf{1}_m, 0) \text{ dan}$$

$$\mathbf{D}_{((m-d+1) \times (m+1))} = \begin{bmatrix} \mathbf{D}_{((m-d) \times m)} & \mathbf{0}_{(m-d)} \\ \mathbf{0}_m^T & 0 \end{bmatrix}$$

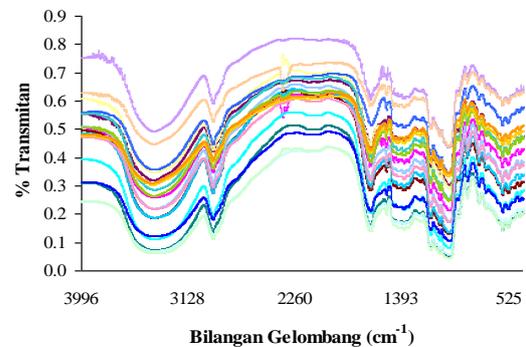
3. Menetapkan ordo pembeda (d) yang digunakan (nilai d bisa divariasikan).
 4. Membangun model RSP pada data kalibrasi, kemudian akan diperoleh nilai dugaan, sisaan dan GCV.
 5. Data bagian kedua digunakan untuk validasi model, kemudian akan nilai konsentrasi dugaan, sisaan residual dan RMSEP.
- C. Memodelkan data yang tidak dikoreksi dan membandingkan hasilnya dengan model pada data yang dikoreksi.

HASIL DAN PEMBAHASAN

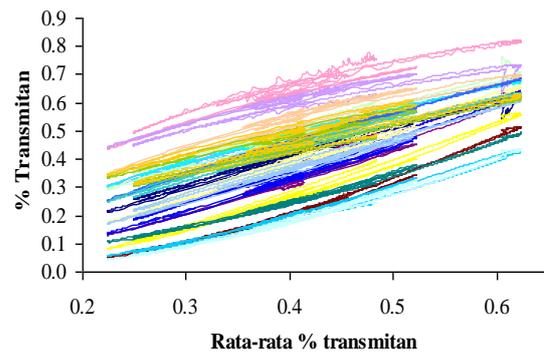
Koreksi Pencaran Multiplikatif

Data persen transmittan merupakan hasil pengukuran menggunakan FTIR pada 1866

bilangan gelombang yang berkisar antara 4000 – 400 cm^{-1} . Grafik spektrum persen transmittan gingerol dari 20 contoh serbuk rimpang jahe disajikan pada Gambar 1. Persen transmittan memiliki interval nilai yang besar antara 0.04 sampai 0.82, sehingga sebaran plot spekturnya berjauhan.



Gambar 1. Spektrum gingerol 20 contoh serbuk rimpang jahe yang tidak dikoreksi

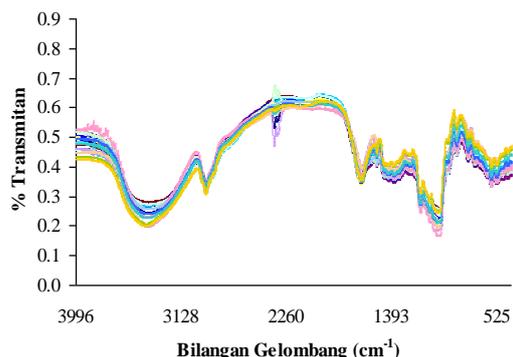


Gambar 2. Plot % transmittan terhadap rata-rata seluruh contoh pada data yang tidak dikoreksi

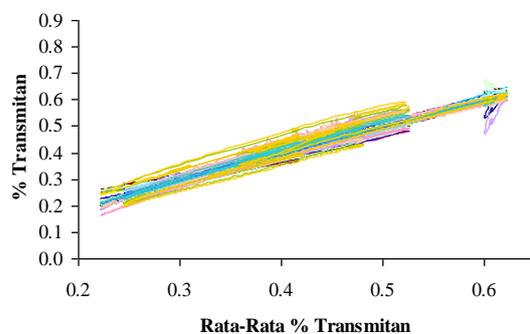
Gambar 2 menunjukkan bahwa ada perbedaan nilai kemiringan dan intercept pada garis regresi spektra contoh terhadap rata-ratanya. Arnita (2005) mengidentifikasi perbedaan tersebut sebagai perbedaan yang diakibatkan adanya pengaruh pencaran. Oleh karena itu, informasi yang ada pada tiap contoh juga berbeda, sehingga perlu dilakukan koreksi pencaran agar masing-masing contoh dapat memberikan informasi yang relatif sama.

Secara berurutan Gambar 3 dan Gambar 4 menyajikan plot spektrum gingerol dan plot persen transmittan terhadap rata-ratanya pada data yang dikoreksi. Gambar 3 memiliki pola yang sama dengan Gambar 2, akan tetapi sebaran plot pada Gambar 4 lebih rapat daripada Gambar 1. Perbedaan nilai kemiringan dan intercept pada Gambar 4 jauh lebih kecil dibandingkan dengan Gambar 2. Hal ini mengindikasikan bahwa informasi yang diberikan oleh seluruh spektrum contoh yang sudah dikoreksi relatif sama. Dengan

demikian, koreksi pencaran multiplikatif mampu mengeliminasi keragaman antar spektra.



Gambar 3. Spektrum gingerol 20 contoh serbuk rimpang jahe yang dikoreksi.



Gambar 4. Plot % transmitan terhadap rata-rata seluruh contoh pada data yang dikoreksi.

Pembentukan Basis B-Spline

Sebelum memodelkan data dengan RSP, terlebih dulu dibangun fungsi basis B-spline. Input yang dibutuhkan dalam membangun B-spline adalah domain, derajat B-spline dan banyaknya *knot* yang digunakan. Domain yang digunakan pada penelitian ini adalah indeks dari 1866 bilangan gelombang yang terukur oleh FTIR, sehingga indeks 1 sampai 1866 dijadikan domain bagi B-spline. Basis B-spline yang digunakan dalam penelitian ini adalah basis B-spline kubik. *Knot* internal yang digunakan untuk membangun B-spline kubik sebanyak 13 *knot* dan 14 *knot* (banyaknya *knot* internal = banyaknya interval *knot* + 1). Penggunaan dua kemungkinan jumlah *knot* bertujuan untuk mendapatkan model yang terbaik. Jumlah *knot* yang banyak menyebabkan *overfitting*, sebaliknya jika terlalu sedikit akan mengakibatkan *underfitting*.

Regresi Sinyal P-Spline

Hasil pendugaan parameter model RSP menggunakan metode Newton-Rapshon sangat dipengaruhi oleh batas bawah dari nilai λ dan κ . Adanya batasan nilai λ dan κ pada fungsi S mengakibatkan dugaan nilai λ dan κ selalu sama

dengan batas bawah yang diberikan. Hal tersebut dapat dilihat dari suku kedua dan ketiga pada Persamaan 6 yang bertanda positif. Oleh karena itu batas bawah yang diberikan untuk λ dan κ bukanlah nol, melainkan bilangan positif yang kecil. Hal tersebut bertujuan untuk mengatasi permasalahan multikolinearitas pada data.

Penentuan nilai batas bawah yang akan digunakan pada penelitian ini didasarkan pada pengalaman. Batas bawah untuk nilai λ dan κ pada data yang dikoreksi masing-masing sebesar 0,015 dan 0,0007, sedangkan pada data yang tidak dikoreksi batas bawahnya adalah 0,0001 untuk λ dan 0,05 untuk κ . Nilai-nilai tersebut digunakan untuk menduga nilai parameter model RSP secara serentak menggunakan metode Newton-Rapshon. Pada data yang dikoreksi, nilai S, GCV dan RMSEP terkecil dicapai pada model RSP yang menggunakan 14 *knot* internal, $d = 1$ dan masing-masing memiliki nilai 0,06246, 0,00004 dan 0,06862. Sedangkan pada data yang tidak dikoreksi, nilai S, GCV dan RMSEP terkecil dicapai pada model RSP yang menggunakan 13 *knot* internal, $d = 1$ dan masing-masing memiliki nilai 0,06481, 0,00004 dan 0,11801. Secara berurutan, Tabel 1 dan Tabel 2 menyajikan ringkasan hasil prediksi model RSP untuk kelompok data kalibrasi dan kelompok data validasi.

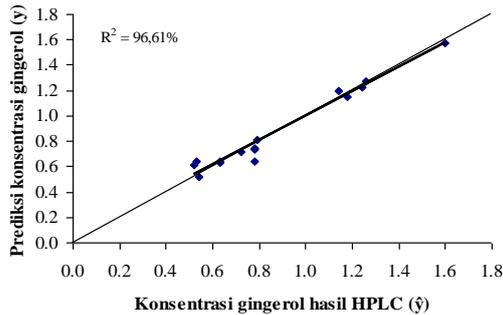
Tabel 1. Nilai y dan \hat{y} konsentrasi gingerol pada kelompok data kalibrasi

Y (HPLC)	\hat{Y} (data dikoreksi)	\hat{Y} (data tidak dikoreksi)
0.53	0.64297	0.64569
0.72	0.71558	0.73211
0.78	0.64297	0.64569
0.52	0.61584	0.62472
0.54	0.51737	0.50782
0.78	0.73290	0.70448
0.63	0.63880	0.63005
0.63	0.63295	0.66354
0.78	0.74834	0.75924
0.79	0.81152	0.78751
1.26	1.27688	1.28782
1.60	1.57313	1.58034
1.18	1.14610	1.15693
1.14	1.20083	1.19486
1.24	1.22257	1.20056

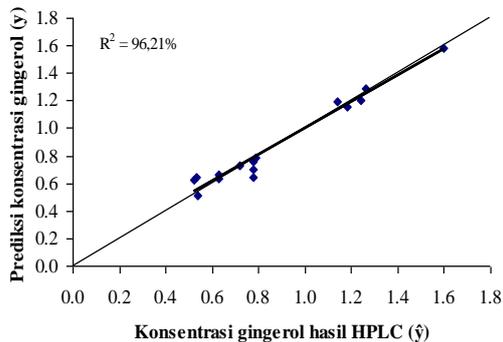
Tabel 2. Nilai y dan \hat{y} konsentrasi gingerol pada kelompok data validasi

Y (HPLC)	\hat{Y} (data dikoreksi)	\hat{Y} (data tidak dikoreksi)
0.63	0.57961	0.59133
0.58	0.41736	0.55498
0.53	0.54331	0.57817
0.79	0.71884	0.75569
1.07	1.25816	1.20388

Nilai R^2 dari regresi \hat{y} terhadap y merupakan salah satu ukuran kebaikan model. Berdasarkan data pada Tabel 2, regresi \hat{y} terhadap y menghasilkan R^2 sebesar 96,61% untuk data yang dikoreksi dan 96,21% untuk data yang tidak dikoreksi. Plot antara nilai y dan \hat{y} untuk kelompok kalibrasi pada Gambar 5a dan Gambar 5b mendekati pola garis lurus 45° yang melalui titik nol. Hal ini menunjukkan bahwa nilai prediksi yang diperoleh RSP cukup dekat dengan nilai y yang sebenarnya untuk kedua kelompok data (data dengan koreksi dan data tanpa koreksi).



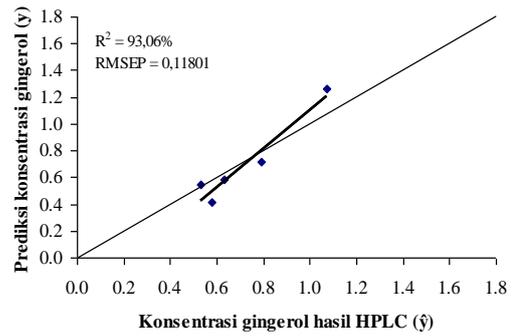
Gambar 5a. Plot y dan \hat{y} pada data kalibrasi yang dikoreksi



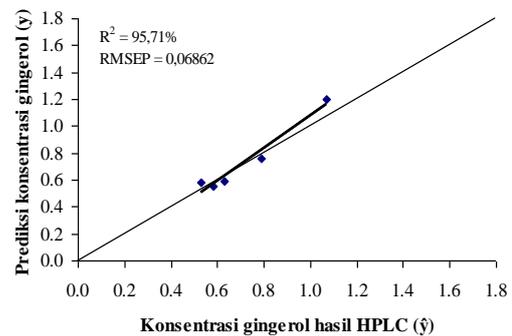
Gambar 5b. Plot y dan \hat{y} pada data kalibrasi yang tidak dikoreksi

Gambar 6a menyajikan plot antara nilai y dan \hat{y} untuk kelompok data validasi yang dikoreksi dan Gambar 6b untuk data validasi yang tidak dikoreksi. Gambar 6a lebih mendekati pola garis lurus 45° yang melalui titik nol dan R^2 yang dimiliki lebih tinggi daripada Gambar 6b. Hal ini berarti bahwa model RSP pada data yang dikoreksi memiliki kemampuan prediksi yang lebih baik dibandingkan dengan pada data yang tidak dikoreksi.

Berdasarkan hasil yang diperoleh, pra-pemrosesan koreksi pencaran pada data perser transmittan senyawa aktif gingerol mampu meningkatkan kemampuan prediksi RSP. Hal ini ditunjukkan dengan adanya penurunan RMSEP sebesar 41,85% dan peningkatan R^2 pada regresi antara nilai y dan \hat{y} .



Gambar 6a. Plot y dan \hat{y} data validasi yang dikoreksi



Gambar 6b. Plot y dan \hat{y} data validasi yang tidak dikoreksi

Hasil yang diperoleh RSP mampu mereduksi RMSEP yang dihasilkan oleh RKU dengan pra-pemrosesan koreksi pencaran (Arnita 2005) sebesar 37,91%. Sedangkan bila dibandingkan dengan hasil dari transformasi wavelet dengan menggunakan RKU sebagai model, RSP mampu mereduksi RMSEP sebesar 35,99% (Sunaryo 2005). RSP ini merupakan pemodelan yang kompetitif bagi pemodelan kalibrasi dengan pendekatan Bayes yang menggunakan regresi terpenggal sebagai teknik pereduksian dimensi data. Gabungan antara pendekatan Bayes dan regresi terpenggal menghasilkan RMSEP sebesar 0,0622 (Erfiani 2005).

Tabel 3. Nilai RMSEP dan $R^2_{y \text{ vs } \hat{y}}$ dari beberapa model kalibrasi pada gingerol

Model	RMSEP	$R^2_{y \text{ vs } \hat{y}}$ (%)
Koreksi pencaran - RKU *	0.10960	82.40
Transformasi wavelet - RKU **	0.10720	93.90
Regresi terpenggal – pendekatan Bayes ***	0.06220	93.90
Koreksi pencaran - RSP	0.06862	95.71

Keterangan : * Arnita 2005; ** Sunaryo 2005; *** Erfiani 2005

SIMPULAN

Koreksi pencaran pada data persen transmitan senyawa aktif gingerol rimpang jahe dapat meningkatkan kemampuan prediksi model regresi sinyal P-spline. Model RSP dengan koreksi pencaran multiplikatif memberikan hasil prediksi yang lebih baik dibandingkan dengan hasil yang diberikan oleh model RCU dengan pra-pemrosesan koreksi pencaran maupun transformasi wavelet. Model regresi sinyal P-spline merupakan alternatif yang baik bagi pemodelan kalibrasi dengan tingkat keakutuan yang cukup tinggi.

DAFTAR PUSTAKA

- [Anonim]. B-Spline Basis Functions: Important Properties. <http://www.cs.mtu.edu/~shene/COURSES/cs3621/NOTES/spline/bsplineproperty.html> [15 Mei 2005].
- Arnita. 2005. Koreksi Pencaran dalam Model Kalibrasi Peubah Ganda pada Data Senyawa Aktif Gingerol Serbuk Rimpang Jahe (*Zingiber Officinale Roscoe*) [tesis]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.
- Atok RM. 2005. Jaringan Syaraf Tiruan dalam Pemodelan Kalibrasi dengan Pra-pemrosesan Analisis Komponen Utama dan Transformasi Fourier Diskret [tesis]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.
- De Boor C. 1978. *A Practical Guide to Spline*. New York: Springer-Verlag.
- Djuraidah A. 2003. Penerapan Model Nonlinear PLS dengan Jaringan Syaraf Tiruan dalam Kalibrasi. *Jurnal Matematika Aplikasi dan Pembelajarannya (JMAP)* 2:339-345.
- Eilers PHC & Marx BD. 1996. Flexible Smoothing with B-Spline and Penalties. *Technometrics* 11:89-121.
- Erfiani. 2005. Pengembangan Model Kalibrasi dengan Pendekatan Bayes (Kasus Tanaman Obat) [disertasi]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.
- Gunawan A. 2001. Studi Penggunaan Pemulusan Spline pada Regresi Nonparametrik [skripsi]. Depok: Fakultas Matematika dan Ilmu Pengetahuan Alam. Universitas Indonesia.
- Hastie T & Tibshirani R. 1990. *Generalized Additive Models*. London: Chapman and Hall.
- Hastie T & Mallows C. 1993. A Discussion of 'A Statistical View of some Chemometrics Regression Tools' by I. E. Frank and J. H. Friedman. *Technometrics* 35:140-143.
- Herwindiati DE. 1997. Pengkajian Regresi Komponen Utama. Regresi Ridge dan Regresi Kuadrat Terkecil Parsial untuk Mengatasi Masalah Kolinearitas [tesis]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.
- Marx BD & Eilers PHC. 1999. Generalized Linear Regression on Sampled Sinyal and Curves: A P-Spline Approach. *Technometrics* 41:1-13.
- Marx BD, Eilers PHC. 2002. Multivariate Calibration Stability: A Comparison of Methods. *J. Chemometrics* 16:129-140.
- Marx BD & Eilers PHC. 2003. Multivariate Calibration with Temperature Interaction using Two-Dimensional Penalized Signal Regression. *Chemometrics and Intelligent Laboratory Systems* 66:159-174.
- Naes T, Issakson T, Fearn T, Davies T. 2002. *A User Friendly Guide to Multivariate Calibration and Classification*. UK: NIR Publications.
- Rahayu W. 2003. Pendekatan Bayes dalam Masalah Kalibrasi [tesis]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.
- Sunaryo S & Notodiputro KA. 2004. Penerapan Metode Transformasi Wavelet Diskret untuk Menentukan Kandungan Senyawa Gingerol pada Tanaman Jahe. *Statistika - Forum Teori dan Aplikasi Statistika* 4: 181-185. Jurusan Statistika FMIPA UNISBA.
- Sunaryo S. 2005. Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-pemrosesan [disertasi]. Bogor: Program Pascasarjana. Institut Pertanian Bogor.