Statistika, Vol. 8 No. 1, 79 – 83
Mei 2008

Beberapa Metode Pendekatan untuk Model Kalibrasi Gingerol

ERFIANI
Departemen Statistika, FMIPA – IPB
Jl. Meranti, Kampus IPB Darmaga, Bogor 16680, Tlp./fax: (0251) 624535
e-mail: efiani@ipb.ac.id, dan Erfiani_ipb@yahoo.com

ABSTRAK

Model Kalibrasi adalah suatu fungsi matematik dengan data empirik dan pengetahuan untuk menduga informasi pada y yang tidak diketahui berdasarkan informasi pada X yang tersedia (Martens dan Naes 1989). Dalam bidang kimia, model kalibrasi merupakan suatu fungsi hubungan antara absorban (X) pada panjang gelombang yang dihasilkan oleh spektrometer dengan konsentrasi (y) larutan unsur atau senyawa yang akan dianalisis (Nur dan Adijuwana, 1989). Dengan kalibrasi, konsentrasi larutan contoh dapat diketahui berdasarkan absorbannya. Beberapa penelitian telah dilakukakan untuk menyusun model kalibrasi Gingerol. Hasil Penelitian menunjukkan dari beberapa pendekatan yang dicobakan diperoleh nila Root Mean Squares Error of Prediction (RMSEP) masing-masing sebagai berikut: Regresi Komponen Utama (0.1096), metode Transformasi Wavelet Diskret (0.1072), Pendekatan Bayes (0.0622), Regresi Sinyal P-spline (0.0686) dan Regresi Kontinum (0.0453). Regresi Kontinum dengan melakukan pre-processing Transformasi Wavelet Diskret ternyata memberikan hasil nilai RMSEP terkecil dan persentase R2y vs $\hat{y}$ terbesar dibandingkan pendekatan lainnya.

Kata Kunci : Model Kalibasi, RMSEP, Regresi Komponen Utama, Transformasi Wavelet Diskret, Pendekatan Bayes, Regresi Sinyal P-spline, Regresi Kontinum

1. Pendahuluan

Model kalibrasi menggambarkan hubungan antara berbagai respons dari instrumen analitik dengan satu atau lebih karakteristik dari suatu bahan aktif. Model ini mengandung parameter yang nilainya harus diduga dari referensi agar dapat digunakan untuk menduga karakteristik dari bahan aktif baru yang belum diketahui. Pada pendugaan model kalibrasi ganda sering timbul masalah kolinearitas di antara peubah absorban (Naes, 1985), sehingga metode baku seperti model regresi sering tidak memberikan solusi yang tidak stabil. Oleh karena itu diperlukan suatu metode yang dapat mengatasi kolinearitas sehingga metode ini akan memberikan solusi yang lebih stabil.

Beberapa metode pendekatan yang dapat digunakan untuk penyusunan model kalibrasi antara lain Regresi Komponen Utama (RKU) dan *Partial Least Square* (PLS) (Draper dan Smith, 1996; Martens dan Naes, 1989; Wold, 1984; Young, 1994). Dalam penyusunan model, RKU menggunakan peubah baru yang merupakan kombinasi linear peubah-peubah asal. Metode PLS menghasilkan komponen-komponen yang tidak berkorelasi atau tidak terjadi kolinearitas tetapi dapat memaksimumkan korelasinya dengan peubah respons (Geladi dan Kowalski, 1986).

Selain kedua pendekatan tersebut terdapat beberapa pendekatan lain yang dapat digunakan untuk menyusun model kalibrasi. Pada kajian ini ingin dilakukan pembandingan penyusunan model kalibrasi untuk senyawa gingerol menggunakan beberapa pendekatan yaitu pendekatan Bayes, metode Transformasi *Wavelet* Diskret (TWD), Regresi Kontinum (RK), RKU, dan Regresi Kuadrat Terkecil Parsial (RKTP).

Data yang digunakan adalah data keluaran FTIR untuk pengukuran senyawa aktif gingerol. Senyawa aktif gingerol tersebut berasal dari 14 contoh tanaman jahe hasil pengamatan dua daerah sentra produksi tanaman obat yaitu Kulonprogo (Jawa Tengah) dan Karanganyar (D.I. Yogyakarta). Pengamatan dilakukan pada periode waktu 27 Juli 2003 sampai dengan 1 Agustus 2003. Data keluaran FTIR yang digunakan berupa 1866 titik persentase transmittan pada interval bilangan gelombang 40 – 400 cm^{-1}.

2. Pendekatan Bayes

Pendekatan Bayes merupakan suatu altematif untuk mengatasi masalah kekolinearan karena dalam pendekatan ini informasi baru ditambahkan ke dalam model dengan cara mengganggap bahwa parameter model berasal dari sebaran tertentu sehingga tidak bersifat deterministik. Sebaran ini dikenal sebagai sebaran prior yang mencerminkan keyakinan kita tentang besamya parameter tersebut. Jika parameter model yang ingin diduga adalah $\underline{\beta}$ dengan $\underline{y}$ sebagai peubah acaknya, maka parameter $\underline{\beta}$ dipilih yang memaksimumkan fungsi kepekatan bersyarat $\pi(\underline{\beta}\,|\,\underline{y})$. Fungsi kepekatan bersyarat ini disebut fungsi posterior.

Secara umum, jika $h(\underline{\beta})$ adalah sebaran prior dari $\underline{\beta}$ dan statistik $\underline{w}= u(y_1, y_2,\ldots,y_p)$ maka sebaran posteriornya adalah:

$$\pi(\underline{\beta}\,|\,\underline{w}) = \frac{f(\underline{\beta},\underline{w})}{g(\underline{w})} = \frac{h(\underline{\beta})f(\underline{w}\,|\,\underline{\beta})}{g(\underline{w})} \tag{1}$$

Dalam hal ini $f(\underline{w}|\underline{\beta})$ adalah fungsi kemungkinan dari $\underline{w}$, $f(\underline{\beta},\underline{w})$adalah fungsi kepekatan bersama dari $\underline{\beta}$ dan $\underline{w}$. Nilai $\underline{\beta}$ dipilih sedemikian sehingga diperoleh $\pi(\underline{\beta}|\underline{w})$. Pemilihan $h(\underline{\beta})$ yang tepat akan dapat memperbaiki fungsi $f(\underline{w}|\underline{\beta})$ sehingga ruang bagi nilai optimum $\underline{\beta}$ menjadi terbatas.

Sebaran prior dapat diperoleh melalui dua pendekatan informative prior atau non informative prior. Pada informative prior, parameter ditetapkan memiliki sebaran tertentu dengan kisaran nilai yang dapat diterima. Sedangkan pada non informative prior, tidak ada informasi tambahan tentang parameter. Pada non informative prior apabila prior bernilai konstan maka penduga parameter dengan metode Bayes akan sama dengan penduga dengan metode kemungkinan maksimum.

Metode lain untuk menentukan sebaran prior yang cukup sederhana dan melibatkan data yang ada adalah dengan metode prior sekawan (Conjugate prior). Prior sekawan ditentukan dengan memeriksa fungsi kemungkinan $L(\beta/y) = f(y/\beta)$, dan memilih sebagai keluarga sekawan adalah kelompok sebaran yang sama dengan fungsi kemungkinannya. Sebagai contoh bila $f(y/\beta)$ menyebar normal maka $L(\beta/y)$ juga akan mengikuti sebaran normal.

3. Transformasi *Wavelet* Diskret (TWD)

Misalkan kita mempunyai data $X = (x_0, x_1, \ldots, x_{p-1})$, dan diasumsikan $p = 2^M$, M bilangan bulat positip. DWT dari x untuk mother wavelet ψ, didefinisikan sebagai

$$d_{j,k} = \sum_{j=0}^{p-1} x(t)\,\psi_{j,k}(t) \tag{2}$$

untuk j = 0, 1, 2, …, M-1 dan k = 0, 1, 2, …, 2j – 1

Vektor data X dapat dihubungkan dengan suatu fungsi f pada interval [0,1) dengan rumusan sebagai berikut:

$$f(t) = \sum_{k=0}^{2^M-1} x_k \; I_{\left\{ k/2^M \le t \le (k+1)/2^M \right\}} \tag{3}$$

Sehingga diperoleh DWT :

$$f(t) = c_{0,0}\;\phi(t) + \sum_{j=0}^{M-1}\;\sum_{k=0}^{2^j-1} d_{j,k}\;\psi_{j,k}(t) \tag{4}$$

dimana $\phi(t)$ merupakan fungsi skala yang disebut father wavelet yang berkaitan dengan bagian pemulusan (smoothing) data. Seperti pada $\psi(t)$, fungsi $\phi(t)$ ini juga dapat membangkitkan basis orthonormal $\{\phi_{j,k}(t)\}$.

Biasanya kita tidak memperhatikan semua level resolusi j, tetapi hanya pada level resolusi tertentu, misalnya J yang berkaitan dengan pengaruh koefisien wavelet yang mendekati nol (threshold), sehingga persamaan (4) dapat ditulis sebagai

$$f(t) = \sum_{k} c_{J,k}\ \phi_{j,k}(t) + \sum_{k=J}^{M-1}\ \sum_{k=0}^{2^j-1} d_{j,k}\ \psi_{j,k}(t) \tag{5}$$

$c_{J,k}$ merupakan koefisien-koefisien pemulusan data. Proses rekonstruksi ulang dari koefisien-koefisien wavelet terhadap pengamatan asli disebut Inverse Discret Wavelet Transform (IDWT). Di dalam wavelet terdapat banyak macam mother wavelet dan metode threshold.

4. Regresi Sinyal P-spline (RSP)

RSP merupakan salah satu pendekatan nonparametrik yang melibatkan penggunaan basis B-sline dan penalti pemulus (penalti pembeda dan penalti ridge) dalam pendugaan parameter koefisien regresi. Dengan memanfaatkan informasi spasial dari peubah penjelas, koefisien regresi yang dihasilkan oleh RSP ada dalam ruang fungsi mulus. Hal ini dicapai dengan cara merepresentasikan koefisien regresi sebagai kombinasi linier dari basis B-sline. Dalam model RSP, indeks peubah penjelas memegang peranan yang sangat penting yaitu untuk mengevaluasi fungsi basis B-spline yang dibangun.

Model kalibrasi dapat dipandang sebagai regresi linear berganda yang dirumuskan:

$$y_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \ldots + \alpha_p x_{pi} + e_i \text{, i = 1, 2, ..., n}$$

$$E\left(\underline{y}_{(nx1)}\right) = \alpha_0 \underline{1}_{(nx1)} + X_{(nxp)} \underline{\alpha}_{(px1)} \tag{6}$$

Matriks X(nxp) memiliki dimensi peubah penjelas yang sangat besar sehingga dimensi vektor koefisien regresi α juga besar. Oleh karena itu dalam proses pendugaan α perlu dilakukan pereduksian dimensi koefisien regresi dengan cara merepresentasikan α sebagai kombinasi linier dari basis B-spline (B) yang memiliki dimensi jauh lebih kecil dari dimensi X.

$$\underline{\alpha}_{(px1)} = B_{(pxm)} \underline{\beta}_{(pxm)} \tag{7}$$

β adalah vektor koefisien basis berdimensi m, m<<<p

Permasalahan mendasar dalam membangun fungsi basis B-spline adalah penentuan jumlah dan penempatan knot yaitu tempat tersambungnya potongan-potongan polinomial pada B-spline. Jumlah knot yang terlalu banyak akan menyebabkan overfitting. Pengoptimalan jumlah dan penempatan knot merupakan masalah linear yang kompleks dan lebih mengarah pada masalah algoritma.

P-spline dapat mengatasi permasalahan diatas dengan cara menentukan terlebih dahulu jumlah knot pada B-spline, dan penempatan knot dilakukan dengan konsep equally spaced knots yaitu mengatur posisi knot sedemikian rupa sehingga jarak antar knot yang satu dengan lainnya sama.

Dengan mensubstitusikan persamaan (7) kedalam persamaan (6) diperoleh

$$E\left(\underline{y}_{(nx1)}\right) = \alpha_0 \underline{1}_{(nx1)} + U_{(nxm)} \underline{\beta}_{(mx1)}$$

$$U_{(nxm)} = X_{(nxp)} \underline{\beta}_{(mx1)} \tag{8}$$

Dengan menggunakan Persamaan (8) masalah yang timbul akibat dimensi peubah yang besar dalam proses pendugaan parameter dapat diatasi tanpa membuang data pada matriks X.

Persamaan (8) memberikan model regresi berganda berdimensi sedang yang terkendala pada suatu ruang fungsi pemulus, akan tetapi masalah multikoliniearitas masih ada pada kolom-kolom matriks U. Multikoliniearitas pada U dapat diatasi dengan menambahkan dua penalti yaitu penalti pembeda dan penalti ridge, karena regresi ridge merupakan salah satu solusi untuk masalah multikolinearitas. Penambahan dua penalti tersebut secara bersamaan dapat meningkatkan kemulusan α.

5. Regresi Kontinum (RK)

Regresi kontinum merupakan salah satu metode yang digunakan untuk mengatasi kolinearitas ganda dan singularitas yang terjadi pada model regresi ganda. Misalkan X adalah matriks data hasil pengamatan yang sudah dipusatkan (centred) berukuran (nxp)dan disebut peubah bebas, sedangkan $\underline{y}$ adalah vektor peubah respon berukuran (nx1) pengamatan yang sudah dipusatkan.

Prinsip Rktp Adalah Memaksimumkan Koragam Antara Peubah Bebas Dengan Peubah Respon. Teknik Rktp Mempunyai Kemiripan Dengan Rku. Perbedaan Penting Kedua Metode Adalah Pada Rku Mengkonstruksi Faktor Yang Dapat Menerangkan Sebanyak-Banyaknya Keragaman Dari Data Peubah Bebas X Tanpa Mempertimbangkan Apakah Faktor Tersebut Berhubungan Dengan Data Peubah Respon $\underline{y}$ Atau Tidak. Di Sisi Lain, Pada Rktp Mengkonstruksi Faktor Yang Mampu Menerangkan Keragaman Data Peubah Bebas X Dan Pada Saat Yang Sama Faktor Tersebut Mempunyai Hubungan Dengan Data Peubah Respon Y.

6. Hasil dan Pembahasan

Erfiani (2005) menyusun model kalibrasi Gingerol menggunakan pendekatan regresi terpenggal pada tahap awal dan pendekatan Bayes dengan perilaku β berhirarki dan σ acak. Model kalibrasi Gingerol dan Kurkumin yang diperoleh memberikan nilai dugaan konsentrasi pengamatan dengan tingkat ketelitian yang cukup tinggi. Besaran *Root Mean Squares Error of Prediction* (RMSEP) yang diperoleh pada model kalibrasi Gingerol adalah 0.06220.

Arnita (2005) menggunakan pre-processing koreksi pencaran multiplikatif dan RKU untuk menyusun model kalibrasi gingerol, dan diperoleh hasil besaran RMSEP sebesar 0.10961.

Sunaryo (2005) menerapkan metode Transformasi *Wavelet* Diskret (TWD) pada data gingerol dan ternyata mampu melakukan reduksi dimensi dengan baik. Meski demikian tidak ada jaminan, bahwa kasus multikolinearitas teratasi. Sehingga TWD sebaiknya digabung dengan metode lain yang mampu mengatasi multikolinearitas dalam pembuatan model kalibrasi peubah ganda. Gabungan wavelet dengan RKU (Regresi Komponen Utama), untuk menduga model prediksi kadar senyawa gingerol pada rimpang jahe, ternyata menghasilkan model yang relatif cukup memuaskan. Menggunakan pendekatan tersebut diperoleh RMSEP untuk model kalibrasi gingerol sebesar 0.10720.

Setiawan (2007) melakukan eksplorasi terhadap tiga metode yaitu Regresi Kontinum (RK), Regresi Komponen Utama (RKU), dan Regresi Kuadrat Terkecil Parsial (RKTP). Aplikasi pendekatan regresi kontinum dengan Transformasi *Wavelet* Diskret (RK-TWD) pada data konsentrasi senyawa gingerol menghasilkan besaran RMSEP lebih kecil dibandingkan dengan hasil pendekatan RKU-TWD maupun RKTP-TWD. Aplikasi pendekatan RK-TWD pada model kalibrasi gingerol memberikan hasil RMSEP sebesar 0.0453.

Tonah (2006), menggunakan pre-processing koreksi pencaran multiplikatif dan RSP untuk menyusun model kalibrasi gingerol, dan diperoleh hasil besaran RMSEP sebesar 0.0.0682.

Setiawan (2007) melakukan eksplorasi terhadap tiga metode yaitu Regresi Kontinum (RK), Regresi Komponen Utama (RKU), dan Regresi Kuadrat Terkecil Parsial (RKTP). Aplikasi pendekatan regresi kontinum dengan transformasi *wavelet* diskret (RK-TWD) pada data konsentrasi senyawa gingerol menghasilkan besaran RMSEP lebih kecil dibandingkan dengan hasil pendekatan RKU-TWD maupun RKTP-TWD. Aplikasi pendekatan RK-TWD pada model kalibrasi gingerol memberikan hasil RMSEP sebesar 0.0453.

7. Kesimpulan

Terdapat beberapa pendekatan yang dapat digunakan untuk menyusun model kalibrasi. Beberapa penelitian telah dilakukan untuk menyusun model kalibrasi gingerol dengan menggunakan beberapa pendekatan. Berdasarkan keseluruhan hasil yang diperoleh model kalibrasi gingerol terbaik diperoleh menggunakan Regresi Kontinum dengan melakukan *pre-processing* Transformasi *Wavelet* Diskret. Ringkasan hasil secara keseluruhan tersaji pada Tabel 1 berikut ini.

Tabel 1. Ringkasan Hasil

METODE	RMSEP	$R^2_{y\ vs\ \hat{y}}$ (%)
Koreksi Pencaran – RKU	0.10961	82.40
Transformasi wavelet – RKU	0.10720	93.90
Regresi terpenggal – Pendekatan Bayes	0.06220	93.90
Koreksi Pencaran – RSP	0.06862	95.71
RK – TWD	0.0453	97.94

Daftar Pustaka

[1]. Arnita. 2005. *Koreksi pencaran dalam model kalibrasi peubah ganda pada data senyawa gingerol serbuk rimpang* (*Zingiber Officinale Roscue*) [Tesis]. Bogor: Program Pascasarjana, Institut Pertanian Bogor.

[2]. Draper, N. Smith, H. 1996. *Applied Regression Analysis*, Second Edition. John Willey & Sons. Chichester.

[3]. Erfiani. 2005. *Model Kalibrasi Gingerol Menggunakan Pendekatan Bayes Hirarki* [disertasi]. Bogor : Sekolah Pascasarjana, Institut Pertanian Bogor.

[4]. Geladi, P and B. R. Kowalski. 1986. Partial least square regression. A tutorial. *Analytica Chimica Acta*, 185: 1-17.

[5]. Martens, H. and T. Naes. 1989. *Multivariate Calibration*. John Willey & Sons. Chichester, England.

[6]. Naes, T. 1985. Multivariate calibration when the error covariance matrix is structured. *Technometrics*, Vol. 27, No. 3: 301-311.

[7]. Nur, M.A. dan H. Adijuwana. 1989. *Teknik Spektroskopi dalam Analisis Biologi*. Pusat antar Universitas Ilmu Hayat, Institut Pertanian Bogor.

[8]. Setiawan. 2007. *Pendekatan Regresi Kontinum dalam Model Kalibrasi* [disertasi]. Bogor : Sekolah Pascasarjana, Institut Pertanian Bogor.

[9]. Sunaryo S. 2005. *Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-pemrosesan* [disertasi]. Bogor : Sekolah Pascasarjana, Institut Pertanian Bogor.

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

[11]. Wold, S. et al. 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM J. SCI. Stat. Comput.* Vol 5, No 3:735-

[12]. Young, P.J. 1994. Reformulation of the partial least square regression algorithm. *Siam J.SCL STAT Comput.*, Vol. 5, No.1: 225-230.