

Penerapan Model Kalibrasi dengan Transformasi Wavelet Diskrit – Partial Least Square (TWD-PLS) Pada Data Gingerol

Application of Caliberation Model using Discrete Wavelet Transformation – Partial Least Square for Giengerol Data

Sony Sunaryo & Sri Mumpuni Retnaningsih
Jurusan Statistika Institut Teknologi Sepuluh November Surabaya

ABSTRACT

The determination of concentration of gingerol compound is usually carried out through a long and expensive process using HPLC instrument. The alternative method to predict such concentration can be done using a multivariate calibration model. Since the numbers of samples (n) are less than the number of variables (p) and between the independent variables are correlated, the development of model using conventional regression is no longer valid. The combination of Discrete Wavelet Transformation (DWT) and Partial Least Square has been adopted in this research to predict concentration of gingerol and it showed a promising result.

Keywords : calibration model, wavelet, PLS, gingerol

PENDAHULUAN

Penentuan kadar senyawa gingerol pada rimpang jahe secara kuantitatif dilakukan melalui proses yang panjang meliputi penghancuran bahan, pelarutan, dan pengukuran dengan HPLC (*High Performance Liquid Chromatography*). Proses tersebut memerlukan waktu dan biaya yang relatif mahal (Naes *et al.* 2002). Pengukuran dengan FTIR (*Fourier Transform Infra Red*) relatif lebih mudah dan murah untuk dilakukan daripada pengukuran dengan HPLC. Setiap bentuk spektrum persen transmittan dari FTIR akan mencerminkan gugus fungsi yang terdapat pada senyawa gingerol dari suatu sampel rimpang jahe. Selain itu intensitas spektra dapat juga digunakan sebagai ukuran secara kuantitatif. Salah satu kelemahan pengukuran data kuantitatif oleh FTIR adalah ketergantungannya terhadap kemurnian sampel yang diukur. Walaupun dengan analisis referensi terhadap senyawa standard, secara umum pola spektrum (pola serapan) yang dihasilkan relatif sama, tetapi jika sampel yang diukur tidak murni maka pengukuran serapan yang terbentuk tidak tajam dan melebar. Hal ini disebabkan untuk sampel yang tidak murni gugus fungsi-gugus fungsi yang sama masih terkandung dalam beberapa senyawa yang berbeda. Untuk sampel yang berupa ekstrak yang murni maka pola serapan yang terbentuk

akan lebih tajam dan lebih spesifik. Sehingga cara alternatif untuk memprediksi kadar gingerol pada rimpang jahe adalah dengan mengembangkan model kalibrasi peubah ganda yang menyatakan hubungan kadar senyawa aktif hasil dari HPLC (sebagai peubah tak bebas Y) dengan data hasil pengukuran bilangan gelombang menggunakan FTIR dari serbuk rimpang jahe yang berupa data spektra persen transmittan (sebagai peubah bebas X).

Kalibrasi peubah ganda adalah disiplin ilmu yang tercakup dalam *chemometrics*, untuk menemukan hubungan antara sekumpulan ukuran yang relatif mudah atau murah memperolehnya dengan sekumpulan ukuran yang memerlukan waktu dan biaya yang relatif mahal (Naes *et al.* 2002). Dari sudut pandang statistika tujuan kalibrasi peubah ganda adalah menemukan model $E(Y) = f(X)$, untuk prediksi Y dengan akurasi dan presisi yang tinggi.

Karena biasanya dimensi peubah bebas X sangat tinggi dan antar peubah saling berkorelasi, maka kasus jumlah pengamatan sampel lebih kecil dari jumlah peubah bebas, dan kasus multikolinearitas sering muncul dalam kalibrasi peubah ganda. Sehingga penanganan dengan metode regresi ganda baku secara langsung kurang valid. Data spektra persen transmittan (X) akan berupa sederetan data vektor $\underline{x} = (x_1, x_2, \dots, x_p)^T$ yang berdimensi tinggi dan saling berkorelasi, sehingga pengembangan model kalibrasi

peubah ganda $E(Y) = f(X)$ dengan mengikutkan semua data X menjadi tidak efisien. Dengan reduksi dimensi, diharapkan pengembangan model kalibrasi peubah ganda menjadi lebih efisien. Reduksi dimensi yang digunakan dalam paper ini adalah metode transformasi *wavelet* diskret (TWD), dengan alasan metode ini setelah dikaji ternyata lebih baik dibanding metode lain seperti transformasi Fourier maupun PCA (*Principal Component Analysis*).

Karena fokus metode *wavelet* yang digunakan hanya sebagai reduksi dimensi, bukan untuk mengatasi kasus multikolinearitas, maka dimungkinkan untuk menggunakan hasil keluaran dari metode *wavelet* sebagai masukan untuk metode kalibrasi peubah ganda yang lain, seperti regresi bertatar, regresi komponen utama (PCR) dan PLS, sehingga akan diperoleh model yang lebih baik. Pada penelitian ini akan dibahas penerapan gabungan metode *wavelet* dan PLS untuk memprediksi kadar senyawa gingerol pada rimpang jahe. Penelitian ini merupakan pengembangan dari penelitian Sunaryo dan Notodiputro (2004, 2005, 2006). Perhitungan matriks koefisien *wavelet* dengan menggunakan *software* wavetresh 3 seperti yang diterangkan oleh Nason (1994, 1998).

METODE

Dari 20 sampel masing-masing untuk serbuk rimpang jahe dan 40 sampel serbuk rimpang temulawak dengan FTIR dihasilkan data spektra persen transmitan yang diamati 1866 titik, pada bilangan gelombang $4000 - 400 \text{ cm}^{-1}$ yang mencerminkan kadar gingerol. Karena jumlah sampel dipandang mencukupi maka sampel-sampel dibagi menjadi 2 kelompok. Untuk rimpang jahe 15 sampel untuk kalibrasi dan 5 sampel untuk validasi, sedangkan untuk temulawak 30 sampel untuk kalibrasi dan 10 sampel sisanya untuk validasi. Pemilihan pada masing-masing kelompok dilakukan secara acak. Karena metode *wavelet* mensyaratkan jumlah titik harus 2^M , untuk M bilangan bulat positif, maka dari 1866 titik diambil 1024 titik dengan memperhatikan daerah identifikasi spektra infra merah gingerol yang memberikan informasi .

Dari 1024 titik yang terpilih dilakukan transformasi *wavelet* diskret (TWD), dengan melihat berbagai kemungkinan resolusi yang menghasilkan koefisien-koefisien *wavelet* yang jumlahnya lebih kecil dari jumlah sampel untuk kelompok data kalibrasi, serta berbagai fungsi mother *wavelet* keluarga Daubechies. Alasan pemilihan mother *wavelet* keluarga Daubechies karena sering dipakai dalam aplikasi dan memberikan hasil pemodelan

yang baik (Brown *et al.* 2001, McNulty & Ganapati 1998, Yi-yu & Chen 2000). Koefisien-koefisien *wavelet* yang dihasilkan digunakan untuk pengembangan model kalibrasi peubah ganda. Perhitungan matriks *wavelet* pada penelitian ini menggunakan *software* wavetresh 3 seperti yang diterangkan oleh Nason (1994, 1998). Karena lama penyimpanan berpengaruh terhadap kadar gingerol yang dihasilkan, maka dalam pencarian model prediksi yang lebih baik diikuti peubah dummy yang mencerminkan kelompok lama penyimpanan (untuk data pada penelitian ini 3 bulan dikode 0 dan 10 bulan dikode 1).

Secara matematis TWD tidak menjamin bahwa antara koefisien-koefisien *wavelet* yang dihasilkan tidak saling berkorelasi, sehingga metode selanjutnya dalam penelitian ini digunakan Partial Least Square (PLS) untuk menghilangkan kasus multikolinearitas yang selanjutnya disebut TWD-PLS. Langkah-langkah analisis dengan metode TWD-PLS dapat dijelaskan sebagai berikut :

Dari data spektrum persen transmitan dapat dituliskan matriks $X_{(n \times p)}$, dimana n adalah banyaknya sampel dan p adalah banyaknya titik persen transmitan yang diteliti pada masing-masing bilangan gelombang. Konsentrasi senyawa aktif

dituliskan dalam $\underline{y}_{(n \times 1)}$. Misalkan matriks X berukuran $(n \times p)$ dan \bar{x} yang terkoreksi terhadap nilai rata-ratanya adalah seperti dinyatakan dalam persamaan (1).

$$\mathcal{X} = X - \underline{1} \bar{x}^T = \begin{bmatrix} \mathcal{X}_{1.}^T \\ \mathcal{X}_{2.}^T \\ \vdots \\ \mathcal{X}_{n.}^T \end{bmatrix} \dots \dots \dots (1)$$

Dengan transformasi *wavelet* diskret

$$\mathcal{X}_{j.}^T = \underline{d}_{j.}^T W$$

dimana W ditentukan oleh mother *wavelet* tertentu, maka akan diperoleh matriks koefisien *wavelet*

$$D_{(n \times p)} = \mathcal{X}_{(n \times p)} W^T_{(p \times p)}$$

Kemudian dengan memilih level-level resolusi tertentu yang jumlah koefisien *wavelet* yang dihasilkan lebih kecil dari $n-1$, maka akan diperoleh

$$D^*_{(n \times m)} = \mathcal{X}_{(n \times p)} W^{*T}_{(p \times m)}$$

yang mereduksi pengamatan dari p titik tiap-tiap sampel menjadi m titik koefisien wavelet yang terpilih.

Persamaan regresi antara $\underline{y}_{(nx1)}$ terhadap $D^*_{(nxm)}$ dapat ditulis seperti dalam persamaan (2) berikut:

$$\underline{y} = \underline{1} q_0 + D^* \underline{q} + \underline{e} \quad (2)$$

dengan q adalah fungsi *mother wavelet* dan q_0 adalah fungsi *father wavelet*. Jika $m < n$ maka dugaan kuadrat terkecil dari persamaan (2) adalah :

$$\hat{\underline{q}} = (D^{*T} D^*)^{-1} D^{*T} \underline{y}$$

Persamaan regresi prediksi linear yang berbentuk seperti persamaan (3):

$$\hat{y}_{pred} = b_0 + b_1 x_1 + \dots + b_p x_p = b_0 + \underline{x}^T \underline{b} \quad (3)$$

dapat dihitung dengan $\underline{b} = W^* \hat{\underline{q}}$ dan $b_0 = \bar{y} - \bar{x}^T \underline{b}$.

Jika multikolinearitas masih terjadi antar koefisien wavelet, maka langkah yang bisa diambil adalah menghitung skor komponen utama dari D^* .

kemudian meregresikan ulang antara $\underline{y}_{(nx1)}$ dengan skor-skor komponen utama. Pemilihan model terbaik dapat dilakukan dengan memperhatikan beberapa ukuran kebaikan model prediksi seperti F , R^2 dan S .

Hasil persamaan \hat{y}_{pred} dalam persamaan (3) akan digunakan untuk memprediksi kadar senyawa aktif kelompok sampel data validasi. Root Mean Square Error of prediction (RMSEP) merupakan salah satu

ukuran yang dapat digunakan sebagai ukuran kebaikan hasil prediksi (McNulty & Ganapati 1998, Shao & Yadong 2004, Yi-yu & Chen 2000) seperti dinyatakan dalam persamaan (4).

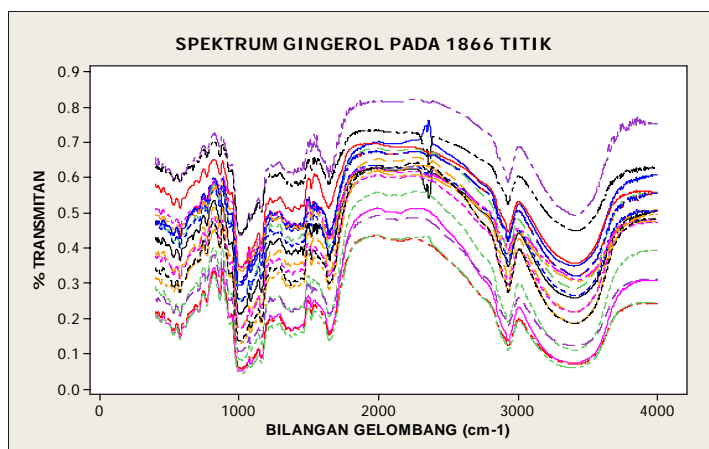
$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_{pred}} (\hat{y}_{i pred} - y_i)^2}{N_{pred}}} \quad (4)$$

dengan N_{pred} adalah banyaknya sampel untuk validasi. Semakin kecil RMSEP, semakin baik prediksi model yang dihasilkan.

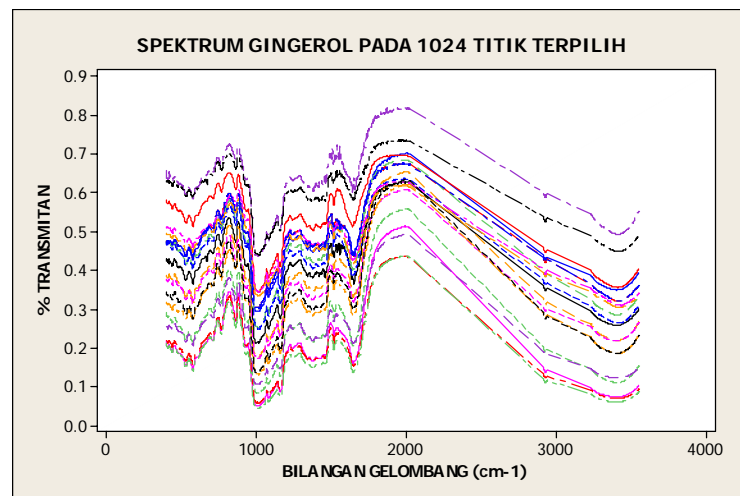
HASIL DAN PEMBAHASAN

Penentuan kadar gingerol

Gambar spektra persen transmittan serbuk jahe pada 1866 titik dan 1024 titik terpilih, untuk 20 sampel serbuk rimpang jahe bisa dilihat pada Gambar 1 dan Gambar 2. Dengan mengambil 11 koefisien wavelet (untuk *mother wavelet* Daubechies - 20) pada resolusi 0, 1 dan 3 serta 1 koefisien untuk fungsi skala hasil transformasi wavelet diskret dilakukan pencarian model terbaik untuk prediksi kadar gingerol. Alasan diambil *mother wavelet* Daubechies - 10 dan level resolusi 0, 1 dan 3, karena perilakunya yang relatif lebih baik dibanding yang lain, dalam arti dapat menangkap ukuran-ukuran kebaikan model seperti R^2 dan S yang relatif lebih baik.



Gambar 1. Spektra persen transmittan 1866 titik, untuk 20 sampel serbuk rimpang jahe.



Gambar 2. Spektra persen transmitan 1024 titik, untuk 20 sampel serbuk rimpang jahe.

Permasalahan yang timbul, ternyata korelasi antar koefisien *wavelet* masih banyak yang tinggi. Sehingga pemodelan dengan regresi ganda biasa dengan peubah respon kadar gingerol dan peubah bebas koefisien *wavelet* menjadi kurang valid, karena masih terjadi kasus multikolinearitas. Hal ini terlihat dari analisis regresi antara kadar gingerol dengan 12 koefisien *wavelet* dan peubah *dummy* waktu penyimpanan diperoleh R^2 yang tinggi (99,9%) sedang semua peubah bebas tidak signifikan, selain itu nilai VIF (*Variance Inflation Factor*) masing-masing peubah berkisar antara

43,4 – 681,2. Metode yang digunakan untuk mengatasi kasus multikolinear antar koefisien *wavelet* dalam penelitian ini adalah PLS. Hasil ukuran-ukuran kebaikan model untuk berbagai kemungkinan kombinasi jumlah komponen dalam PLS dapat dilihat pada Tabel 1. Dari Tabel 1 ternyata dengan PLS 5 komponen pertama dari 12 komponen serta satu peubah *dummy* menghasilkan model yang relatif lebih baik dibandingkan yang lainnya. Sehingga model inilah yang dipilih untuk model prediksi kadar gingerol.

Tabel 1. Ringkasan Nilai Kebaikan model Gingerol dengan *wavelet* D-10 dan PLS.

Banyak Komponen PLS (+ dummy)	F	S	R^2	RMSEP
1	41,93**	0,02803	76,3336%	0,265972
1- 2	31,47**	0,020547	83,9865%	0,256195
1-3	41,91**	0,011260	91,9556%	0,230614
1-4	56,89**	0,006481	95,7906%	0,131669
1-5	65,78**	0,004557	97,3365%	0,111938
1-6	75,33**	0,003347	98,2608%	0,116571
1-7	116,39**	0,001874	99,1481%	0,158265
1-8	154,80**	0,001237	99,5179%	0,207364
1-9	194,14**	0,000879	99,7146%	0,218533
1-10	218,74**	0,000703	99,8175%	0,133292
1-11	185,49**	0,000753	99,8532%	0,140410
1-12	115,64**	0,001108	99,8561%	0,149496

Keterangan : ** signifikan pada $\alpha = 0.01$, F: nilai F_{hitung} uji model regresi, S : Kuadrat Tengah Galat

Plot Y dengan \hat{Y} untuk kelompok kalibrasi dapat dilihat pada Gambar 3. Gambar 3 menunjukkan bahwa model kalibrasi kadar gingerol untuk menduga 15 sampel kelompok data kalibrasi cukup memuaskan. Hasil dari TWD-PLS digunakan untuk menduga 5 sampel kelompok data validasi. Ringkasan prediksi untuk kelompok sampel data kalibrasi dan kelompok sampel data validasi model TWD-PLS dapat dilihat pada Tabel 2.

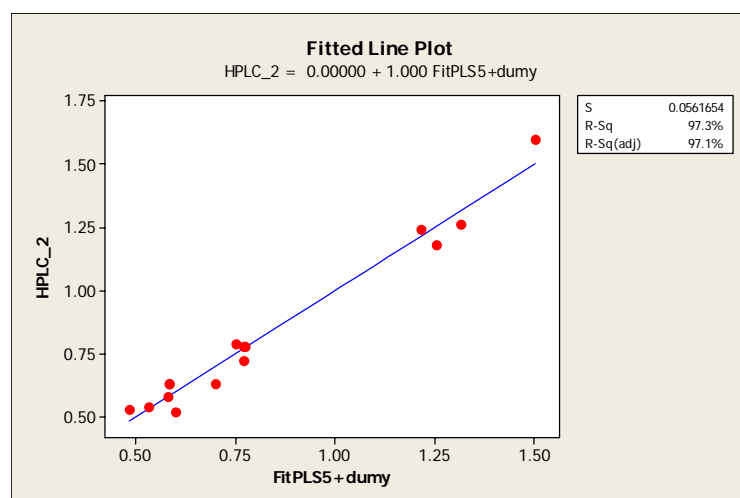
Sedangkan plot antara Y dengan \hat{Y} untuk kelompok sampel data validasi dapat dilihat pada Gambar 3 dan Gambar 4. Gambar 4 menunjukkan titik-titik yang ada dekat dengan garis lurus, sehingga model untuk prediksi data

eksternal (yaitu data yang tidak diikutkan dalam pemodelan) cukup memuaskan. Sehingga dapat disimpulkan pada penelitian ini diperoleh model untuk prediksi gingerol dengan $R^2 = 97,3365\%$ dan $RMSEP = 0,111938$.

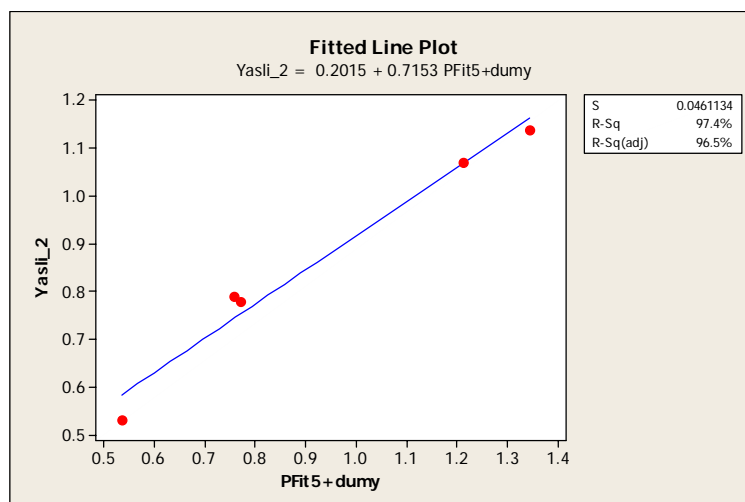
Hasil penelitian untuk data yang sama yang diolah oleh Sunaryo (2006) dengan TWD-PCR, yang menghasilkan $R^2 = 96,7\%$ dan $RMSEP = 0,1072$. Sedangkan dengan pra-pemrosesan *multiplicative scatter corrections* (MSC) menghasilkan $R^2 = 94,6\%$; $RMSEP = 0,1096$. Sehingga hasil analisis dengan TWR-PLS menunjukkan hasil yang cukup memuaskan dan bisa bersaing dengan metode yang lain.

Tabel 2. Nilai Y dan \hat{Y} kadar gingerol dengan *wavelet* D10-PLS (1-5)

KELOMPOK DATA KALIBRASI		KELOMPOK DATA VALIDASI	
Kadar Gingerol dari HPLC (%)	Dugaan (%)	Kadar Gingerol dari HPLC (%)	Dugaan (%)
0,63	0,70069	0,53	0,53458
0,72	0,77036	0,78	0,77012
0,58	0,57917	0,79	0,75713
0,53	0,48279	1,14	1,34350
0,52	0,59942	1,07	1,21156
0,54	0,53061		
0,79	0,75108		
0,78	0,77012		
0,63	0,58242		
0,63	0,58138		
0,78	0,77465		
1,26	1,31557		
1,60	1,50263		
1,18	1,25359		
1,24	1,21552		



Gambar 3. Plot Y dengan \hat{Y} kelompok data kalibrasi gingerol.



Gambar 4. Plot Y dengan \hat{Y} kelompok data validasi gingerol (RMSEP =0,111938)

KESIMPULAN

Metode Transformasi *Wavelet* Diskret (TWD) mampu melakukan reduksi dimensi dengan baik, tetapi tidak ada jaminan, bahwa kasus multikolinearitas teratasi. TWD sebaiknya digabung dengan metode lain yang mampu mengatasi multikolinearitas dalam pembuatan model kalibrasi peubah ganda.

Gabungan Transformasi *wavelet* Diskrit (TWD) dengan *Partial Least Square* (PLS), untuk menduga model prediksi kadar senyawa gingerol pada rimpang jahe, ternyata menghasilkan model yang cukup memuaskan.

DAFTAR PUSTAKA

Brown PJ, Fearn T, Vanucci M. 2001. Bayesian Wavelet Regression on Curves with Application to a Spectroscopic Calibration Problem. *J Amer Statist Assoc* **96**:398-408.

McNulty SC, Ganapati M. 1998. *Application of Wavelet Analysis Determining Glucose Concentration of Aqueous Solution Using NIR Spectroscopy*. Hewlett-Packard comp.

Naes T, Isaksoon T, Fearn T, Davies T. 2002. *A User Friendly Guide to Multivariate Calibration and Classification*. NIR publications, UK.

Nason GP, Silverman BW. 1994. The discrete wavelet transform in S. *J.comp graph. Stat.* **3**:163-191.

Nason GP. 1998. *Wavethresh 3 software*. Department of Mathematics, University of Bristol, UK. (<http://www.stats.bris.ac.uk/~wavethresh>) [20 juni 2003]

Shao X & Zhuang Y. 2004. Determining of Chlorogenic Acid in Plant Samples by Using Near-Infrared Spectrum with Wavelet Transform Preprocessing. *Analytical Sciences* **20**.

Sunaryo S & Notodiputro KA. 2004. Penerapan Metode Transformasi Wavelet Diskret untuk Menentukan Kadar Senyawa Gingerol pada Rimpang Jahe. *Statistika - Forum Teori dan Aplikasi Statistika* **4**:181-185 , Jurusan Statistika FMIPA UNISBA.

Sunaryo S & Notodiputro KA. 2005. Penerapan Metode Transformasi Wavelet Diskret untuk Menentukan Kadar Senyawa Kurkuminoid pada Rimpang Temulawak. *Prosiding Seminar Nasional Matematika*:100-107, Jurusan Matematika UNS, Surakarta, 7 Mei 2005.

Sunaryo S. 2006. Model Kalibrasi dengan Transformasi Wavelet sebagai Metode Pra-Pemrosesan [Disertasi] . Bogor: Program Pascasarjana, Institut Pertanian Bogor

Yi-yu C & Chen M. 2000. A New Computing Multivariate Spectral Analysis Method Based on Wavelet Transform. *Journal of Zhejiang University Science* **1**:15-19.