

5-1-2014

Ridge Regression in Calibration Models with Symmetric Padding Extension-Daubechies Wavelet Transform Preprocessing

Nurwiani

STKIP PGRI Jombang, Indonesia, nurw_13iem64@yahoo.com

S Sunaryo

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Setiawan

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

B W. Otok

Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Nurwiani; Sunaryo, S; Setiawan; and Otok, B W. (2014) "Ridge Regression in Calibration Models with Symmetric Padding Extension-Daubechies Wavelet Transform Preprocessing," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 1 , Article 16.
DOI: 10.22237/jmasm/1398917700

Ridge Regression in Calibration Models with Symmetric Padding Extension-Daubechies Wavelet Transform Preprocessing

Nurwiani

STKIP PGRI Jombang
Jombang, Indonesia

S. Sunaryo

Institut Teknologi
Sepuluh Nopember
Surabaya, Indonesia

Setiawan

Institut Teknologi
Sepuluh Nopember
Surabaya, Indonesia

B. W. Otok

Institut Teknologi
Sepuluh Nopember
Surabaya, Indonesia

Wavelet transformation is commonly used in calibration models as a preprocessing step. This preprocessing does not involve all results of a spectrum discretization; consequently, a lot of information can be missing. To avoid missing information, a symmetric padding extension (SPE) can be used to place all data points into dyadic scales, however, high dimensional discretization points need to be reduced. Dimension reduction can be performed with Daubechies wavelet transformation (DWT). Scale function and Daubechies wavelet are continuous functions, thus they perform a faster approximation. SPE-DWT preprocessing combines SPE and DWT. Multicollinearity often occurs in calibration models; the ridge regression (RR) method can be used to solve multicollinearity problems. This article proposes the RR method with SPE-DWT preprocessing. The proposed method is applied to determine a model for predicting the content of curcumin in turmeric. Selection of the best model is carried out by comparing coefficient of determinations, p-values of the Kolmogorov-Smirnov (KS) error models, and Root Mean Square Error Prediction (RMSEP). Results show that the RR method with SPE-DWT preprocessing gives an accurate prediction.

Keywords: Calibration models, Daubechies wavelet transform, symmetric padding extension

Introduction

In calibration models, the number of observations is usually much smaller than the number of points resulted from the spectrum discretization obtained from Fourier Transform Infrared (FTIR). In preprocessing calibration models, some researchers use Daubechies wavelet transformation (DWT) without involving all

Nurwiani is a lecturer in the Mathematics Department. Email her at: nurw_13iem64@yahoo.com. Sunaryo, Setiawan and Otok are promotors in the Statistics Department.

RIDGE REGRESSION IN CALIBRATION MODELS

points resulted from discretization; this means that a lot of information will be missing in building models. Using DWT, Brown et al. (2001) involved 2^8 of 700 points in building models to estimate the content of fat, sugar, flour and water in bread dough. Using Haar wavelet transform, Sunaryo and Retnaningsih (2008) used 2^{10} of 1,866 points to estimate the content of gingerol in ginger. In these two studies there are as many as 444 and 842 missing information points, respectively.

To avoid missing information, symmetric padding extension (SPE) can be used with all data points in dyadic scales. Spectrum discretization points are predictor variables with an original size of p , changed to $q = 2^M$, where M is a positive integer and $q \geq p$. Dimension reduction can be performed with DWT. The scale function and Daubechies wavelet are continuous functions; thus, they can perform a faster approximation.

To date, SPE-DWT preprocessing, which combines SPE and DWT in calibration models, has not been used. SPE-DWT preprocessing avoids information loss during preprocess and determines the orthogonal matrix in dimension reduction process. Multicollinearity often occurs in calibration models; the ridge regression (RR) method can be used to solve multicollinearity problem. This manuscript proposes a RR method with SPE-DWT preprocessing.

Methodology

Several methods are available to categorize discretization points of wavelength into dyadic scales; one method is the SPE. According to Boggess & Narcowich (2001), SPE is defined as a spectrum that is evenly extended at the endpoints by reflection in two ways: (i) discretization points are reflected around mid-line between the end point and the next point, expressed by SPE1, and (ii) discretization points are reflected in the line through the two end points, expressed with SPE2. Figures 1(a) and 1(b) illustrate the SPE1 and SPE2 of 10 discretization points.

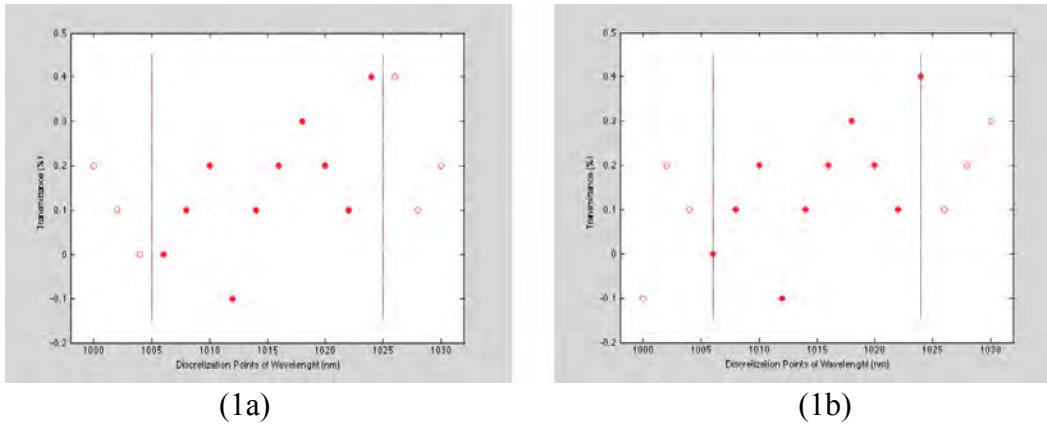


Figure 1: SPE1 (1a) and SPE2 (1b) of 10 Discretization Points

If the discretization points are matrix $\mathbf{X}_{n_1 \times p}^*$, then centering and SPE are performed on this matrix, the resulting matrix will be $\mathbf{X}_{n_1 \times q}^*$. The next step is determining the orthogonal matrix, size $q \times q$, for dimension reduction process by using the wavelet analysis.

There are two main functions in wavelet analysis: scale function ϕ (father wavelet) and wavelet function ψ (mother wavelet). Both functions produce a family of functions that can be used to solve or reconstruct a spectrum (Boggess & Narcowich, 2001). Daubechies wavelet (Daubechies, 1992) is one of these wavelet functions. The scale function and Daubechies wavelet are continuous functions and, thus, can perform a faster approximation. Advantages of the Daubechies wavelet are compact support (closed and bounded), and that the width of support depends on the number of vanishing moments L (which limits the pedestal width) (Daubechies, 1992).

A smoother scaling function and the Daubechies wavelet function can be determined by choosing the power $\ell = 2L - 1$ and filter length $N = 2L$. In the Daubechies wavelet, for each L number of vanishing moments, there will be $2L$ coefficient scales with non-zero values. The scale and Daubechies wavelet functions are located on the interval $0 \leq t \leq 2L - 1$. Daubechies wavelet is commonly expressed by dbN for $L = 2, \dots, 10$ or by $db2L$ for $L = 2, \dots, 10$ (Burrus et al., 1998; Boggess & Narcowich, 2001). To define the Daubechies wavelets, consider the two functions $\phi(t)$ and $\psi(t)$ which are solutions to the following equations:

RIDGE REGRESSION IN CALIBRATION MODELS

equation of scale function,

$$\phi(t) = \sqrt{2} \sum_{k=0}^{N-1} h_k \phi(2t-k), \quad (1)$$

and equation of wavelet function,

$$\psi(t) = \sqrt{2} \sum_{k=0}^{N-1} g_k \phi(2t-k). \quad (2)$$

Filter coefficients of the scale function $\phi(t)$ in (1) for the Daubechies wavelet must satisfy:

$$\sum_{k=0}^{N-1} h_k = \sqrt{2} \quad (3)$$

$$\sum_{k=0}^{N-1} (-1)^k k^m h_k = 0, \quad m = 0, 1, \dots, \frac{N}{2} - 1 \quad (4)$$

$$\sum_{k=0}^{N-1} h_k h_{k+2m} = 0, \quad m \neq 0 \quad (5)$$

$$\sum_{k=0}^{N-1} h_k^2 = 1 \quad (6)$$

A set sequence of scale filter numbers $H = \{h_0, h_1, h_2, \dots, h_{N-1}\}$ called low-pass filters in a pyramid algorithm can be obtained from (3), (4), (5) and (6) (see Burrus et al. (1998) for a detailed discussion about the pyramid). The relationship between the scale filter coefficients h_k and wavelet filter coefficients g_k :

$$g_k = (-1)^k h_{N-1-k} \quad (7)$$

A set sequence of wavelet filter numbers $G = \{g_0, g_1, \dots, g_{N-1}\}$ called high-pass filters in the pyramid algorithm is obtained from (7); based on the orthogonal wavelet matrix that satisfies (1) and (2) it can be determined using:

$$(\mathbf{W}^*)^t \mathbf{W}^* = \mathbf{W}^* (\mathbf{W}^*)^t = \mathbf{I}_q. \quad (8)$$

After centering and SPE are performed, the matrix \mathbf{W}^{*t} in (8) is used for dimension reduction. Dimension reduction can be done by determining the diagonal of matrix $\Lambda_{q \times q}^*$ which are the eigenvalues of symmetric matrix $(\mathbf{X}'\mathbf{X})_{q \times q}$ (Anton & Rorres, 2005, p566, eq1) Hence, the diagonal of matrix $\Lambda_{q \times q}^*$ is obtained from:

$$(\mathbf{W}^*)'_{q \times q} (\mathbf{X}'\mathbf{X})_{q \times q} (\mathbf{W}^*)_{q \times q} = \Lambda_{q \times q}^*.$$

Considering the proportion of

$$p(\lambda_r^*) = (\lambda_1^* + \lambda_2^* + \dots + \lambda_r^*) / (\lambda_1^* + \lambda_2^* + \dots + \lambda_q^*),$$

matrix $\mathbf{W}_{q \times r}^t$ for dimension reduction is obtained, where $r < q$. Dimension reduction of predictor variables and parameters is determined with:

$$\mathbf{Z}_{n_1 \times r} = \mathbf{X}_{n_1 \times q} \mathbf{W}_{q \times r}^t \text{ and } \boldsymbol{\gamma}_{r \times 1} = \mathbf{W}_{r \times q} \boldsymbol{\beta}_{q \times 1}$$

Because the points of the spectrum discretization resulted in calibration models are generally highly correlated, it is necessary to carry out multicollinearity detection. According to Shi-ji & Zhi-bin (1993), by taking into account the type of condition number

$$\text{cond}(\mathbf{Z}'\mathbf{Z}) = \frac{\lambda_{\max}}{\lambda_{\min}},$$

multicollinearity can be detected through:

- i. If $0 < \text{cond}(\mathbf{Z}'\mathbf{Z}) < 100$, there is no multicollinearity (Type I);
- ii. If $100 \leq \text{cond}(\mathbf{Z}'\mathbf{Z}) \leq 1000$, there is some moderate or stronger degree of multicollinearity (Type II);
- iii. If $\text{cond}(\mathbf{Z}'\mathbf{Z}) > 1000$, there is some serious degree of multicollinearity (Type III).

In this study, the active compound curcumin in turmeric is the response variable determined from extraction using High Performance Liquid Chromatography (HPLC). Because data for this response variable does not follow a normal distribution, Johnson Transformation (JT) is carried out. The original

RIDGE REGRESSION IN CALIBRATION MODELS

response variable is $\mathbf{y}_{n_1 \times 1}^*$, whereas $\mathbf{y}_{n_1 \times 1}$ is the response variable that has been transformed and centered. Given the normal distribution of y^* :

$$f(y^*) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y^* - \mu)^2 / 2\sigma^2},$$

for

$$-\infty < \mu < \infty, \sigma > 0, -\infty < y^* < \infty.$$

If y^* is not normally distributed, the JT can be determined using:

$$f(y^*) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}, \quad -\infty < v < \infty, \quad -\infty < y^* < \infty$$

where $v = \alpha + \eta f_i(y^*, \delta, \xi)$, $i = 1, 2, 3$, v is a standard normal random variable, α and η are the shape parameters, δ is a scale parameter and ξ is a location parameter. It is assumed that $\eta > 0$ and $\delta > 0$. Based on its curve, the JT can be differentiated into three systems (George, 2007):

- i. v_{S_b} with f_1 Bounded system, $f_1(y^*, \delta, \xi) = \ln\left(\frac{y^* - \xi}{\delta + \xi - y^*}\right)$, $\xi \leq y^* \leq \delta + \xi$
- ii. v_{S_l} with f_2 Log-normal system and $f_2(y^*, \delta, \xi) = \ln\left(\frac{y^* - \xi}{\delta}\right)$, $y^* > \xi$
- iii. v_{S_u} with f_3 Unbounded system, $f_3(y^*, \delta, \xi) = \sinh^{-1}\left(\frac{y^* - \xi}{\delta}\right)$, $-\infty < y^* < \infty$

where: $\sinh^{-1}\left(\frac{y^* - \xi}{\delta}\right) = \ln\left(\left(\frac{y^* - \xi}{\delta}\right) + \sqrt{1 + \left(\frac{y^* - \xi}{\delta}\right)^2}\right)$. After the dimensional

reduction process, calibration models are obtained as (Naes et al., 2002):

$$\mathbf{X}_{n_1 \times q} = \mathbf{Z}_{n_1 \times r} \mathbf{W}_{r \times q} + \mathbf{F}_{n_1 \times q}, \quad (9)$$

$$\mathbf{y}_{n_1 \times 1} = \mathbf{Z}_{n_1 \times r} \boldsymbol{\gamma}_{r \times 1} + \boldsymbol{\varepsilon}_{n_1 \times 1}. \quad (10)$$

The eigenvalues

$$(\mathbf{Z}'\mathbf{Z})_{r \times r} = \boldsymbol{\Lambda}_{r \times r} = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_r)$$

can be obtained from (9), where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r$ and $\lambda_i > 0, i=1, 2, \dots, r$. The parameter estimation using ordinary least square can be defined as:

$$\hat{\boldsymbol{\gamma}}_{r \times 1} = \boldsymbol{\Lambda}_{r \times r}^{-1} \mathbf{Z}'_{r \times n_1} \mathbf{y}_{n_1 \times 1}$$

$$\text{where } \hat{\mathbf{Z}}_{n_1 \times r} = \mathbf{X}_{n_1 \times q} \hat{\mathbf{W}}'_{q \times r}, \quad \hat{\boldsymbol{\beta}}_{q \times 1} = \mathbf{W}'_{q \times r} \hat{\boldsymbol{\gamma}}_{r \times 1},$$

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \quad \text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{Z}'\mathbf{Z})_{r \times r}^{-1} \text{ and}$$

$$\text{MSE}(\hat{\boldsymbol{\beta}}) = \sigma^2 \sum_{i=1}^r \lambda_i^{-1}.$$

One of the methods to overcome the multicollinearity problem is the RR method (Hoerl & Kennard, 1970). From (10) ridge parameter estimation can be obtained as:

$$\hat{\boldsymbol{\gamma}}_{q \times 1}(\boldsymbol{\theta}) = (\mathbf{Z}'\mathbf{Z} + \boldsymbol{\theta}\mathbf{I})_{q \times q}^{-1} \mathbf{Z}'_{q \times n_1} \mathbf{y}_{q \times 1}$$

where $\theta > 0$, and the ridge parameter can be determined using k iterations (see Shi-ji & Zhi-bin (1993) for a detailed explanation of RR).

In the process of building models it is important to validate the selected regression models. According to Neter et al. (1989), the regression models can be validated by dividing the data into two parts ($n = n_1 + n_2$). The first set of data n_1 , called the model building set, is used to build the model. The second data set n_2 , called the validation or prediction set, is applied to validate the model.

Results

Calibration models using the RR method with SPE-DWT preprocessing are built to predict the content of curcumin in turmeric. The discretization points resulted from FTIR spectrum and the content of curcumin in turmeric determined by HPLC is shown in Figure 2. Figure 2(a) shows the FTIR spectrum of a sample with $p=1866$ points, Figure 2(b) demonstrates the 616 first discretization points of the first sample data, Figure 2(c) illustrates the FTIR spectrum of 40 samples and Figure 2(d) shows the percentage of curcumin in turmeric as determined by the HPLC from 40 samples. These 40 samples are then divided into two parts, the first part of the data set consists of $n_1 = 30$ samples as a model building set and the second part of data set comprises of, $n_2 = 10$ samples as a prediction set.

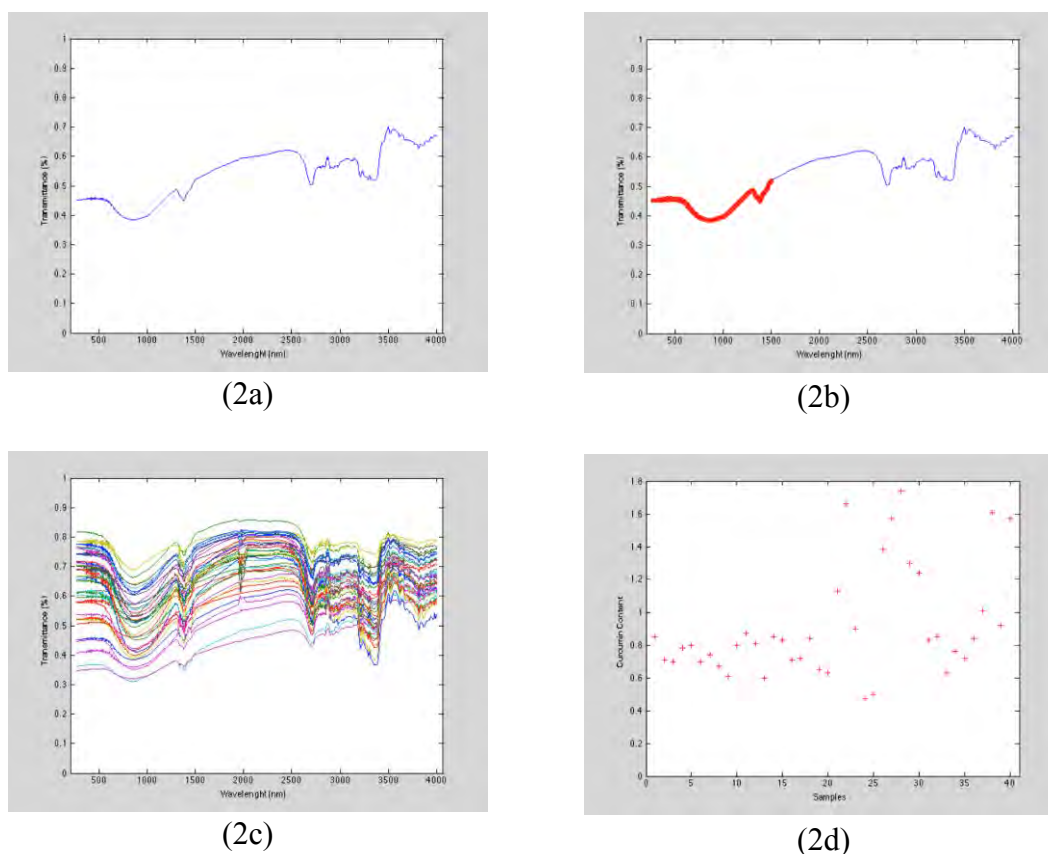


Figure 2. Data of Curcumin in Turmeric

In general, the number of samples in the calibration models is limited, therefore, it is important to conduct normality test for the response variables. There are 30 observations of curcumin in turmeric that do not meet the normality assumption as shown with the Kolmogorov-Smirnov (KS) p value = $3.7478e-13$ (see Figure 3(a)). The JT on the response variable yields:

$$v_{S_U} = -0.568871 + 0.784968 \sinh^{-1} \left(\frac{y^* - 0.708465}{0.0817548} \right)$$

with the KS p value = 0.6320. Further, y , a centered value of v_{S_U} , is defined and illustrated in Figure 3(b). In this study, the building set is carried out only for SPE1 on predictor variables, thus, the number of discretization points is $q = 2^{11}$.

Using DWT through the pyramid algorithm of h_k and g_k for $N=10$, an orthogonal matrix \mathbf{W}_{2048}^* is obtained. In data processing, the RR method requires $r < n_1 - 1$ where $n_1 = 30$. Hence dimension reduction is done by determining the number of transformation matrices for $r=1, 2, \dots, 28$, and finally, the reduced matrix $\hat{\mathbf{W}}_{2048 \times r}^t$ is obtained.

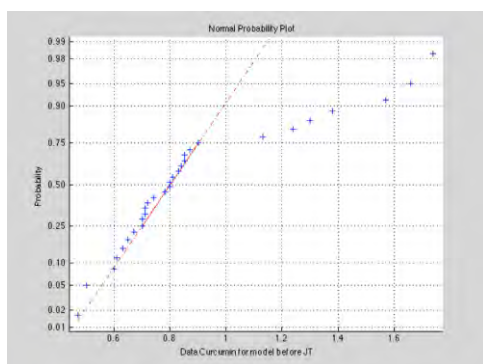
For $r=1$ and $r=2$, the reduced matrix $\hat{\mathbf{W}}_{2048 \times r}^t$ yields $\text{cond}(\mathbf{Z}'\mathbf{Z})$ of Type I, while for $r=3$, it yields $\text{cond}(\mathbf{Z}'\mathbf{Z})$ of Type II. For $r=4, 5, \dots, 28$ the reduced matrix $\hat{\mathbf{W}}_{2048 \times r}^t$ yields $\text{cond}(\mathbf{Z}'\mathbf{Z})$ of Type III.

The RR method to overcome the multicollinearity problem among predictor variables is completed with multiple iterations. Table 1 presents the results of this study. The best model gives the coefficient of determination (Johnson & Whincern, 2002; Seber & Lee, 2003) $R^2 > 0.85$, has the smallest root mean square error (RMSE) (Naes et al., 2002) and has a KS p value error model more than 0.05 (Marsaglia et al., 2003). As Table 1 shows, the RR method with the SPE-DWT preprocessing can be used to build the best models for accurate prediction.

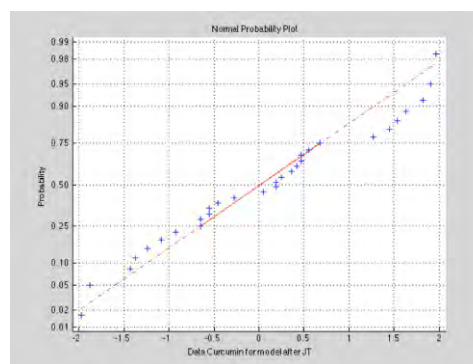
RIDGE REGRESSION IN CALIBRATION MODELS

Table 1: SPE1-Wavelet Daubechies Ridge Regression Models

SPE1	Iteration	KS <i>p</i> value error model	R^2	RMSEP
$r = 1$ $\text{cond}(\mathbf{Z}'\mathbf{Z}) = 1$ $p(\lambda_1^*) = 83.9266$	12,000	0.3333	85.8942	0.4513
	12,100	0.3123	91.4549	0.4551
	12,200	0.2922	97.5712	0.4592
	12,225	0.2873	99.1949	0.4603
	12,230	0.2864	99.5245	0.4605
	12,235	0.2854	99.8556	0.4607
$r = 2$ $p(\lambda_2^*) = 90.0857$ $\text{cond}(\mathbf{Z}'\mathbf{Z}) = 23.3715$	11,650	0.3337	85.4124	0.4518
	11,700	0.3228	88.1813	0.4537
	11,800	0.3017	94.1391	0.4578
	11,850	0.2915	97.3455	0.4599
	11,875	0.2865	99.0096	0.4610
	11,885	0.2846	99.6871	0.4615
$r = 3$ $p(\lambda_3^*) = 92.8664$ $\text{cond}(\mathbf{Z}'\mathbf{Z}) = 901.7588$	8,625	0.3828	85.3135	0.3258
	8,650	0.3727	87.2917	0.3261
	8,700	0.3517	91.4624	0.3268
	8,725	0.3416	93.6611	0.3272
	8,750	0.3316	95.9399	0.3276
	8,775	0.3218	98.3023	0.3280
	8,785	0.3179	99.2715	0.3282
	8,790	0.3160	99.7615	0.3283
	8,791	0.3156	99.8599	0.3283
	8,792	0.3152	99.9584	0.3283



(3a)



(3b)

Figure 3: Normal Probability Plot of 30 Curcumin Data

Conclusion

In calibration models, response variables often do not meet the normal distribution assumption; therefore, the JT is necessary to fulfil model assumptions. The SPE-DWT with filter 10 is able to reduce the dimension, however, there is no guarantee that it can cope with multicollinearity problem. An effective method is needed to overcome the multicollinearity problem. This study shows that the combination of JT and SPE-DWT preprocessing in the RR method can be used to build models that will give accurate predictions. Further study is suggested by implementing the RR method with determination of optimum ridge parameter.

References

- Anton, H., & Rorres, C. (2005). *Elementary linear algebra*. New York: John Wiley & Sons, Inc.
- Boggess, A., & Narcowich, F. J. (2001). *A first course in wavelets with fourier analysis*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Brown, P. J., Fearn, T., & Vannucci, M. (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem. *Journal of the American Statistical Association*, 96: 398-408.
- Burrus, C. S., Gopinath, R. A., & Guo, H. (1998). *Introduction to wavelets and wavelet transforms*. Upper Saddle River, NJ: Prentice-Hall, Inc.
- Daubechies, I. (1992). *Ten lectures on wavelets* (5th ed.). Philadelphia, PA: Society for Industrial and Applied Mathematics.
- George, F. (2007). *Johnson's system of distributions and microarray data analysis*. (Unpublished doctoral dissertation). University of South Florida, Tampa, FL.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12: 55-67.
- Johnson, R. A., & Winchurn, D.W. (2002) *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Marsaglia, G., Tsang, W., & Wang, J. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software*, 8(18): 1-4.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). *A user-friendly guide to Multivariate Calibration and Classification*. Chichester, UK: IM Publications.

RIDGE REGRESSION IN CALIBRATION MODELS

Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models* (2nd ed.). Homewood, IL: Irwin, Inc.

Seber, G. A., & Lee, A. J. (2003). *Linear regression analysis*. New York: John Wiley & Sons.

Shi-ji, C., & Zhi-bin, Z. (1993). Generalized multivariate ridge regression estimate and criteria $Q(c)$ for choosing matrix K . *Applied Mathematics and Mechanics*, 14: 73-84.

Sunaryo, S., & Retnaningsih, S. M. (2008). The implementation of calibration models with discrete wavelet transformation-partial least squares (TWD-PLS) on gingerol data (in Indonesian language). *Ilmu Dasar*, 9: 51-56