Vandana Publications
IJEMR

Volume-7, Issue-2, March-April 2017
International Journal of Engineering and Management Research
Page Number: 166-170

Comparison of Mother Wavelet for Continuous Wavelet Transformation in Multivariate Calibration Model (Case Study of Gingerol Concentration Data)

Dewi Pujo Ningsih[1], Erfiani[2], Aji Hamim Wigena[3]
[1,2,3]Department of Statistics, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, INDONESIA

ABSTRACT

The wavelet transformation is a technique that represents a curve as a combination of basis functions. These are obtained by dilation and translation of a mother wavelet. If the dilation and translation use continuous numbers,they are called as Continuous Wavelet Transformation (CWT). The results of its transformation depend on the mother wavelet type. Mother wavelet Mexican Hat is very good in illustrating the properties of Continuous Wavelet Transformation. However, another mother wavelet often used in the application is Daubechies. This study compared the mother wavelet Daubechies and Mexican Hat in multivariate calibration modeling by using gingerol data. The results showed that the mother wavelet Daubechies was better than Mexican Hat.

***Keyword*—** continuous wavelet transformation, multivariate calibration, principal component regression, dimension reduction

I. INTRODUCTION

The wavelet transformation is a mathematical technique used as a signal processing in analytical chemistry since 1989. The wavelet transformation uses Multire solution Analysis (MRA) techniques to overcome the weakness of previous methods. This technique analyzes the signal at different frequencies with different resolutions. Multire solution Analysis is designed to give good time but poor frequency resolution at low frequencies,and, poor time but good frequencies resolution at high frequency. The basic idea is that the wavelet transformation method represents a curve as a linear combination of basis functions [1].The basis functions obtained by dilation and translation wavelet function are called as mother wavelet ψ. If the dilation and translation use continuous number, theyare called as Continuous Wavelet Transformation (CWT). Meanwhile, if the number of dilation and translational is an integer not negative, it is included in Discrete Wavelet Transformation (DWT)[2]. Compared to the DWT, CWT is a more redundant transformation. It tends to reinforce the traits of its redundancy. Therefore, CWT has much more capability of extracting subtle information from seriously overlapping signals. Furthermore, in CWT, the scale parameter a and translation parameter b vary continuously; we can choose the exact value of a to depict component of a certain frequency band. [3].

The purpose of multivariate calibration model is to find relationship between a set of measurement that is relatively easy or cheap to acquire with a set of measurement that is relatively requires much time and money. Problems often encountered in multivariate calibration modeling are very high dimension of independent variables X and inter-correlated variables. The consequence of the condition in which number of observations is smaller than the number of variables is ununiquemodeling parameter estimation. Reduction dimension is one of alternative ways to overcome this problem. There are several methods of dimension reduction that can be applied, such as Fourier Transformation and Principal Component Analysis. The Fourier transformation uses cosine basis functions, while the principal component analysis forms the new variables that maximizes the variation among variables. Another dimension reduction method is Wavelet Transformation which is better than Fourier Transformation and Principal Component Analysis [1].

In applying wavelet transformation, it is necessary to select mother wavelet in order to establish basis functions. Therefore, the results of the analysis wavelet transformation depend on mother wavelet that has been chosen. According to [4], mother wavelet Mexican Hat is very good in illustrating the properties of CWT. Another mother wavelet that is often used in the application and provides good modeling is Daubechies[1].

This study aims to compare mother wavelet Mexican Hat and Daubechies in multivariate calibration modeling. It uses gingeroldata research in 2003-2005 for estimating active compound gingerol in ginger plants [5]. The common method used to determine the levels of the active compounds from medicinal plants is HPLC (High Performance Liquid Chromatography) which through the long process. Another alternative method is multivariate calibration by modeling concentration value using HPLC method and transmittance percent data through FTIR method which is more efficient in time and cost. Statistically, transmittance percentageof FTIR method has enormous dimensions where n<<p and each variables are correlated. The modeling has been done by using Principal Component Regression (PCR) and CWT. As a pre-processing for dimension reduction using CWT, Mexican Hat and Daubechies are chosen as the mother wavelet. In the end, the estimation result of these two mother wavelet will be compared.

II. DATA AND OBJECTIVE OF STUDY

This study uses secondary data with 20 samples. The data consists of transmittance percent ages as independent variable (X) and gingerolconcentration as dependent variable (y). Transmittance percentage data has been discretized as much as 1866 points on each sample, so that the independent variable's dimensions are 20x1866. The number of the independent variables which is more than the sample number leads to un unique parameter estimation. One method to overcome this problem is to reduce the dimension of independent variables. Another problem that often arises in multivariate calibration modeling is multi colinearity. Therefore, CWT method using Dau bechies and Mexican Hat mother wavelet, combined with the Principal Component Analysis. The results of these two are compared based on the value of RMSEP.

III. METHODOLOGY

There are two main problems in multivariate calibration modeling. First, the number of independent variables is more than the dependent variable, and multicolinearity. To resolving this problems, analysis uses following steps:

Data Transformation

Independent variables $\mathbf{x_1}$, $\mathbf{x_2}$, ..$\mathbf{x_{20}}$are transformed with CWT using the formula

$$w(a,b) = \int_{-\infty}^{+\infty} x(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt\,, a > 0$$

Where$a$ and $b$are scale and translation parameter, $x(t)$are signals that have been analyzed, and$\psi(t)$is mother wavelet function.

The transformation applied Daubechies and Mexican Hat as mother wavelet on 1,2, …, 6 scale each. This transformation is resulting CWT's coefficient matrix which has the same size with the origin variable i.e. 20x1866.

Dimension Reduction of Independent Variable (X)

The transformation is resulting CWT's coefficient matrix which are sized 20x1866. This absolute coefficients value are sorted and the nineteen biggest value are taken on each sample. Next, choose CWT's coefficient which are accommodated by at least 75% of total sample. Only CWT's coefficient fulfilling the criteria will be modeled. So, we obtain new independent variable of CWT's coefficient sized 20x19.

Modeling with Ordinary Least Square (OLS) Regression

Dependent variable y isgingerol concentration data which is sized 20x1 and independent variable of wavelet coefficients (20x19) is modeled by using OLS regression.

Multicolinearity checking

Multicolinearity checking using formula $\left(VIF\right)_i = \frac{1}{1-R_i^2}$,where R_i^2 is the determination coefficient of the ith-explanatory variables, are regressed against the other explanatory variables. If the VIF<10, there is no multikolinear indication (Draper & Smith 1981). Otherwise, if it is indicated multicollinearity, Principal Component Analysis (PCA) is applied to this data.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is performed to determine how many principal components that meet the cumulative variation proportion of the origin variable that can be explained by the i-th principal components.

The i-th principal component; Wi which is formed based on variables that have been standardized in which $Z' = (Z_1, Z_2, \dots, Z_p)$ with cov($\mathbf{Z}$) = $\mathbf{\Sigma}$defined as follows:

$$W_i = e_{i1}Z_1 + e_{i2}Z_2 + \cdots + e_{ip}Z_p \qquad i = 1, 2, \dots, p$$

Meanwhile, the proportions of total variance that can be explained by the k-th principal components based on standardized independent variables are defined as follows:

$$\begin{pmatrix} Total\ of\ population\ variation's\ proportion \\ that\ explained\ by\ k-th\ principal\ component \end{pmatrix}$$

$$= \lambda\_k/(tr(\mathbf{\Sigma})) = \frac{\lambda_k}{p}$$

where$\lambda_k$is eigenvalue of$\mathbf{\Sigma}$ and$k$=*1, 2, …, p*

Modeling using Principal Component Regression (PCR)

Modeling is applied to the active compound concentration data $\mathbf{y}$ (20x1) and the selected CWT's coefficient X (20x19). Next, the data for modeling and data for validation is determined by using re-sampling techniques without replacement, using proportion 3: 1. At each iteration, modeling training data is uses PCR. Furthermore calculate the value of root mean square error of prediction (RMSEP) using the formula $RMSEP=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}$. The smaller value, the smaller difference between the estimated value to the actual value, which means the obtained model is more accurated in generating estimation value. Then calculating the average from 1000 iteration.

Determination of the mother wavelet that gives better prediction

At final stage, average of RMSEP of two mother wavelet at each scaleis obtained. This average then, is compared to find out which one is the best based on the smallest RMSEP.

IV. PRIOR APPROACH

Some methods to overcome problems of large variable's dimension and multicolinearity problem in building multivariate calibration model have been applied to the gingerol data research. Atok (2005) used Principal Component Analysis and DiscreteFourier Transformation with a combination of Artificial Neural Networks, Sunaryo (2005) usedDiscrete Wavelet Transformation (DWT) and Principal Component Analysis (PCA) with RMSEP 0.1072 & 0.1715, Setiawan (2007) used DWT approach and Continuum Regression (CR) with RMSEP 0.0453 & 0.0867. The obtained results showed that wavelet transformation as preprocessing method gave betterresult.

Wavelet expansions and wavelet transforms had been proven to be very efficient and effective in analyzing a very wide class of signals and phenomena [5]. One of its properties that give this effectiveness was because the wavelet expansion allowed more accurate local description and separation of signal characteristics. Fourier coefficient represented a component that lasts for all time and, therefore, temporary events must be described by a phase characteristic that allowed cancellation or reinforcement over large time periods. A Wavelet expansion coefficient represented a local component that waseasier to interpret.

Prior approach used discrete wavelet transforms\ (DWT). In the wavelet method, if the dilation and translation use continuous numbers, they are called as Continuous Wavelet Transformation (CWT). Consequently, during the computation, wavelet analysis can be shifted subtly over the whole area of the original signal which was analyzed by any scale. Meanwhile, if the number of dilation and translational an integer, it was included in DWT.

Discrete Wavelet Transformation (DWT) requiredthe data observation point in multiple 2. The discrete wavelet transformation method was about how to represents time and scale of a signal which was using digital filtering technique and sub sampling operation

V. OUR APPROACH

Fisrt step of analysis, the independent variable (X) is transformed byusing CWT method. Independent variable (X) is transmittance percentagedatafrom 20samples. Transmittance percentage of each sample is discretized by 1866 observation point, so the matrix X's size is 20x1866. The transformation of independent variables is computerized by Matlab software with Mexican Hat as the mother wavelet. It is chosen because this mother wavelet is very good in illustrating the properties CWT[4]. Transformation is startedfrom taking a = 1 and CWT can be calculated for all values of$a$. On the other hand, according to [6], depending on the signal, complete transformation is usually unnecessary or, in other words, transformation calculation for the limited interval is usually appropriate. Therefore this study is using the value of a = 1,2,3,4,5 and 6 only.

Continuous Wavelet Transformation's coefficients are the result of inner product signals to the mother wavelet. Wavelet coefficients shows the correlation between mother wavelet and the signal, so wavelet coefficients will be greater if mother wavelet fits the signal. This CWT's coefficients will become the independent variable. will be The CWT coefficient matrix has the same size with the origin variable i.e. 20x1866. The number of independent variable which are more than observation number leads to not unique model parameters prediction, therefore it will apply reduction dimension of wavelet coefficient.

Dimension reduction is done by sorting the absolute value of the coefficient CWT descending. Nineteen largest value of each sample are recorded by its p-th's number of variable.However, it should be noted that inselecting variables, all p between the sample 1 into sample all 20 should coincide. Therefore, the sample proportion that accommodated each of the CWT's coefficient needs to be checked. This study is limited into 75%, which means there are at least 15 samples that accommodated the chosen CWT's coefficient.

Another mother wavelet wanted to be compared in this study is Daubechies10. The reason for choosing it because it is often used in the application and provide good modeling [1]. The similar steps also applied as mother wavelet Mexican Hat. In Figure 1, we can see the proportion of samples which accommodates the nineteen largest CWT's coefficient on the Daubechies 10 and Mexican Hat with scale 1.2, .., 6.In Figure 1 (a)can be observed on the Mexican Hat, each line represents a scale. From the results obtained, it is only mother wavelet scale of 5-6 giving the sample proportion for the 19 largest coefficient CWT ≥ 75%. Therefore, the next to be modeled is the mother wavelet coefficients CWT of

Mexican Hat scale 5 and 6. The election of 75% limit is taken with consideration of adjusting the proportion of data modeling. By choosing variable that accommodate all/ majority of sample, it is expected to provide a modeling which is better in representing the relationship between concentration and transmittance percentage.

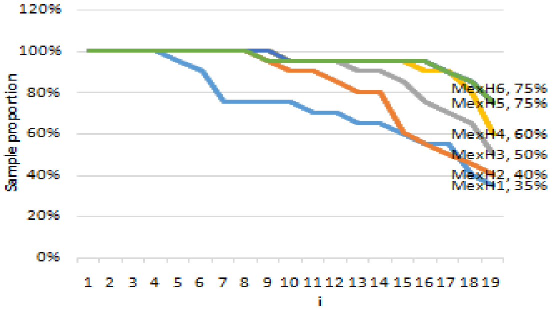

Figure 1. i-th largest CWT's coefficientMother Wavelet Mexican Hat scale 1,2,.., 6

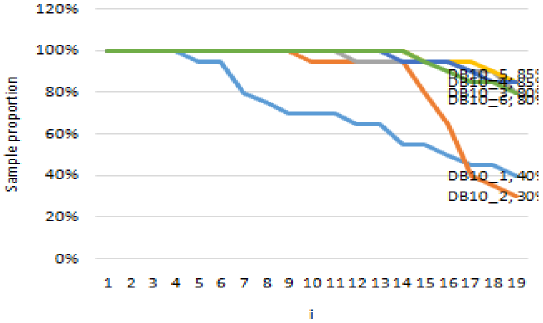

Figure2. i-th largest CWT's coefficientMother Wavelet Daubechies scale 1,2,.., 6

Multicoliearity checking gives VIF > 10, therefore PRC is applied and gives cumulative percent variances of principal component as follows:

TABLE I
CUMULATIVE PERCENT VARIANCES

i-th's PC	DB10_3	DB10_4	DB10_5	DB10_6	MEXH5	MEXH6
1	90.15	90.92	91.09	91.81	91.03	90.99
2	96.34	96.52	97.15	97.27	96.95	96.91
3	99.08	99.75	99.66	99.94	99.97	99.99
4	99.88	99.97	99.94	99.98	100	100
5	99.98	99.99	99.99	100	100	100
6	99.99	100	100	100	100	100
7	100	100	100	100	100	100

With 1000 times iteration, the obtained average and standard deviation RMSEP values for various k-combinations principal components used in the modeling are presented in Table 2. The smaller value of RMSEP means the closer estimated value concentration of the active compound which is derived from the model compared to the real value.

TABLE II
RMSEP

nk	DB10_3		DB10_4		DB10_5		DB10_6		MEXH5		MEXH6	
	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s	$\bar{x}$	s
1	0.311	0.086	0.314	0.086	0.313	0.087	0.315	0.086	0.313	0.087	0.313	0.089
2	0.306	0.092	0.346	0.098	0.336	0.19	0.347	0.098	0.339	0.097	0.34	0.1
3	0.209	0.065	0.355	0.115	0.336	0.262	0.353	0.103	0.356	0.107	0.345	0.111
4	0.227	0.066	0.387	0.117	0.36	0.371	0.388	0.124	0.371	0.116	0.374	0.449
5	0.254	0.076	0.422	0.149	0.404	0.642	0.42	0.135	0.368	0.114	0.379	0.664
6	0.271	0.085	0.458	0.166	0.397	1.249	0.401	0.131	0.428	0.157	0.411	0.664
7	0.317	0.117	0.482	0.156	0.411	0.099	0.423	0.142	0.472	0.169	0.424	0.418
8	0.342	0.121	0.554	0.182	0.453	0.1	0.44	0.144	0.585	0.184	0.475	0.31
9	0.372	0.138	0.604	0.221	0.494	0.114	0.358	0.138	0.628	0.198	0.364	0.419
10	0.452	0.178	0.645	0.258	0.503	0.128	0.364	0.15	0.634	0.216	0.363	0.209
11	0.429	0.21	0.735	0.316	0.565	0.12	0.44	0.25	0.674	0.23	0.406	0.481
12	0.512	0.299	0.844	0.4	0.672	0.124	0.557	0.339	0.681	0.244	0.548	1.027
13	0.79	0.549	0.918	0.693	0.869	0.132	0.879	0.604	0.783	0.377	0.896	0.917
14	1.216	0.926	1.488	1.481	1.429	0.138	1.274	1.018	0.848	0.507	1.116	1.334

nk : number of principal component, DB: mother wavelet Daubechies, MEXH: mother wavelet Mexican Hat, $\bar{x}$: mean , s : standard deviation

From Table II above can be seen on Daubechies 10 scale of 4-6 and Mexican Hatscale5-6, smallest average of RMSEP obtained from a model that uses one principal component. However, the best RMSEP was obtained from mother wavelet Daubechies 10 scale 3 using three principal components. This model has the cumulative percent of variances 99.08%, average of RMSEP 0.20910 and the standard deviation of the RMSE 0.065.

VI. CONCLUSION

Based on results can be concluded:

1. Mother wavelet Daubechies gives better results than the Mexican Hat
2. Modeling calibration using DWT-PCR produce RMSEP at 0.1072, while the method of CWT- PCR 0.20910. DWT-PCR method better than CWT- PCR based on the value RMSEP.

REFERENCES

[1] Sunaryo. S, Model KalibrasidenganTransformasi*Wavelet*SebagaiMetodePra-Pemrosesan [disertasi], Bogor(ID): InstitutPertanian Bogor, 2005.

[2] Nason GP, Silverman BW. 1997. Wavelets for Regression and other Statistical Problems. Bristol (UK): School of Mathematics University of Bristol.

Fearn T. 1999. Data Compression : FT or *Wavelet*. Spectroscopy Europe, London (http://195.173.150.81/td_col.html).

[3] Chau F, Yi Ze L, Junbin G, Xue-Guang S, Chemometrics From Basics to Wavelet Transform, New Jersey (US): A John Wiley & Sons Inc, 2004.

[4] Addison. P, The Illustrated Wavelet Transform Handbook, London (UK): Institute of Physics Publishing, 2002.

[5] Erfiani, PengembanganModel KalibrasidenganPendekatan Bayes [disertasi], Bogor(ID): InstitutPertanian Bogor, 2005.

[6] Polikar, R. The Wavelet Tutorial 2nd ed. (users.rowan.edu/~**polikar**/***WAVELETS*****)**