Available online at SciVarse ScienceDirect

Chinese Herbal Medicines (CHM)

ISSN 1674-6384

Journal homepage: www.tiprpress.com E-mail: chm@tiprpress.com

Original article

Concentration Prediction of Total Flavonoids in *Aurantii Fructus* Extraction Process: Locally Weighted Regression versus Kinetic Model Equation Based on Ficks Law

Yang Chen[1], Jun-hui Shen[2], Jian Ni[1*], Meng-jie Xu[1], Hao-ran Dou[1], Jing Fu[1], Xiao-xu Dong[1]

1. *School of Chinese Materia Medica, Beijing University of Chinese Medicine, Beijing 100102, China*
2. *Information Centre, Beijing University of Chinese Medicine, Beijing 100029, China*

ARTICLE INFO

Article history

Received:

Revised:

Accepted:

Available online:

DOI:

ABSTRACT

Objective To predict the total flavonoids concentration of *Aurantii Fructus* fried with bran in its extraction process. **Methods** Ultraviolet spectrophotometry was used to determine the concentration of total flavonoids in different extraction time (t) and solvent load (M). Then the predicted procedure was carried out using the following data: 1) based on Ficks second law, the parameters of the kinetic model could be deduced and the equation was established; 2) Locally weighted regression (LWR) code was developed in the WEKA software environment to predict the concentration. And then we used both methods to predict the concentration of total flavonoids in new experiments. **Results** After comparing the predicted results with the experimental data, the LWR model had better accuracy and performance in the prediction. **Conclusion** LWR is applied to analyze the extraction process of Chinese herb for the first time, and it is totally fit for the extraction. LWR-based system is a more simple and accurate way to predict than the established equation. It is a good choice especially for a process which exists no clearly rules, and can be used in the real-time control during the process.

Key words

Aurantii Fructus; kinetic model; locally weighted regression; total flavonoids prediction

1. Introduction

Extraction is the most crucial process that affects the production in Chinese pharmaceutical industry. However nowadays we faced the several challenges for deciding the extraction processes, such as long-term exploration, low-efficiency, no uniform standard, difficulty in prompt analysis during the extraction after the change of process's parameters (Mai et al, 2014). In the recent studies, orthogonal design (Zhang et al, 2013b), uniform design (Wu et al, 2011), star point design (Chen et al, 2014), etc are applied to multiple factors-levels investigation, nevertheless the factors or levels are fixed and need to re-conduct the test if they changed (Bezerra and Santelli, 2008; Firlbeck et al, 2013). In order to solve this problem, we would like to conduct the experiments to study the extraction process in the method of water reflux extraction which is the most widely used method of Chinese medicine production. And then the best alternative was established to predict the concentration of index when the factors and levels changed. *Aurantii Fructus* fried with bran

[*]**Corresponding author: Ni J** Tel: +86-10-8473 8607 E-mail: njtcm@263.net

Fund: National Nature Science Foundation of China (surface project) (81173563)

was chosen in our experiment as an example.

Several studies have been reported on the usage of physical-chemistry theories to describe the extraction process of Chinese materia medica (CMM) (Spiro and Selwood, 1982; Li and Cheng, 1997; Hou et al, 2000; Siepmann and Siepmann, 2008; Su et al, 2011). Most of the kinetic models developed in these papers were based on Fick's first and second laws. However many of the models can not fit the Chinese patent medicine (CPM) production, because the abnormal usage of extracting methods and solvent in those studies. With the reference to the practice, our group deduced the model based on Fick's second law that could be applied to the water reflux extraction, which was the most widely used method in Chinese pharmaceutical industry. And we already tested the accuracy of the model (Han et al, 2011; Zhang et al, 2013a).

To establish the equation of the mathematical model, there are few parameters that we need to determine, such as water absorption coefficient (R) and total flavonoid content (related to f), which is time-consuming and may reduce the accuracy. The linear relationship of natural logarithm of total flavonoids concentration to the extraction time and solvent volume was confined to such a range and it can not accurately predict if the levels exceed outside the range (This may relate to the saturation of solvent). So we selected the locally weighted regression (LWR) which could take a short time and predict without limit. It has been proved to be successfully applied in many fields (Wang and Liu, 2012; Mei et al, 2001). However, this regression algorithm is never introduced in the process studied now.

2. Materials and methods

2.1 Materials and reagents

Tu180 UV Spectrophotometer was from Beijing Purkinje General Instrument Co., Ltd.; ZDHW Electric Sets were from Beijing Zhongxing Albert Instrument Co., Ltd. Naringin standard substance was from Yuexu Material Technology Co., Ltd., Shanghai, China (Batch No. 10236-47-2). Bran fried *Aurantii Fructus* was purchased from Beijing Tongrentang Co., Ltd. and was identified as unripe fruit of *Citrus aurantium* L. by Prof. Chun-sheng Liu in Beijing University of Chinese Medicine.

2.2 Model theory explanation

The mathematical model and LWR were developed with different theories and both of them are interesting and easy to understand.

2.2.1 Mathematical model based on mass transfer theory and Fick's law

Over the years, many kinetic models have been established to describe the extraction process of components from CMM (Spiro and Selwood, 1982). The most popular models applied in natural medicines are based on the mass transfer theory and Fick's first or second law of diffusion. The extraction of natural medicines, actually, is the process of mass transfer of the compounds. And there are two different fileds of mass transfer process such as convective mass transfer and molecular diffusion, and natural medicines extraction belongs to the later. So Fick's first or second law of diffusion is introduced. Then based on these two theories, our group developed the kinetic model which was best for industrial production. As reported (Han et al, 2011; Zhang et al, 2013a), the equation is as follows.

$$C_B = [\frac{a\mathrm{f_1}}{\sigma_1(M-R)}\mathrm{t}^{\frac{1}{2}}]^{\frac{1}{1-\mathrm{n}}} \quad (1\text{-}1)$$

If we used herb pieces, the σ_1 is fixed and we can merge it into parameter α, and (1-1) becomes 1-2.

$$C_B = [\frac{\alpha\mathrm{f_1}}{\mathrm{M}-\mathrm{R}}\mathrm{t}^{\frac{1}{2}}]^{\frac{1}{1-n}} \quad (1\text{-}2)$$

The functional relation of factors and compound concentration is deduced in Table 1.

Table 1 Functional relation of factors and compound concentration

Factors	Equations	
extraction time (t)	$\ln C_B = \frac{1}{1-n}\ln\frac{af_1}{M-R} + \frac{1}{2(1-n)}\ln t$	(1-3)
solvent load (M) / (mL·g^{-1})	$\ln C_B = \frac{1}{1-n}\ln af_1 t^{\frac{1}{2}} - \frac{1}{(1-n)}\ln(M-R)$	(1-4)

C_B: concentration of the compound (mg/mL); M: solvent load (mL/g); t: exaction time (min); σ_1: particle size (μm); f_1: fixed and related to ingredient content in herb; R: water absorption coefficient (μm·g/mL)

For *Aurantii Fructus*, σ_1, f_1, and R are fixed. So if we choose the different levels of factors (M and t), we can get a C_B by using the determination methods established. All we need to do is to determine the parameters n and a according to the functional relation between the factors and C_B, and then the equation can be deduced and we can use the equation to do the prediction when the level of factor changes.

2.2.2 LWR

LWR (Eibe et al, 2003; Ian, 2006) generates the local models at prediction time by giving higher weight to the instances in the neighborhood of the particular test instance and performs a regression on the weighted data.

Training instances close to the test instance receive a high weight, those far away receive a low one. In other words, a linear instance model is tailor-made for the particular test instance at hand and used to predict the instance's value (Cost and Salzberg, 1993). LWR is a memory-based method that performs a regression around a point of interest using only training data "local" to that point. The points used as training data are chosen using a kernel. Kernel based learning methods are a class of statistical learning algorithms. It is proved to be quiet useful to deal with nonlinear structure by reducing nonlinear algorithms to algorithms that are linear in some feature spaces. The kernel shape is a designed parameter for which there are many possible choices: Figure 1 shows a Gaussian

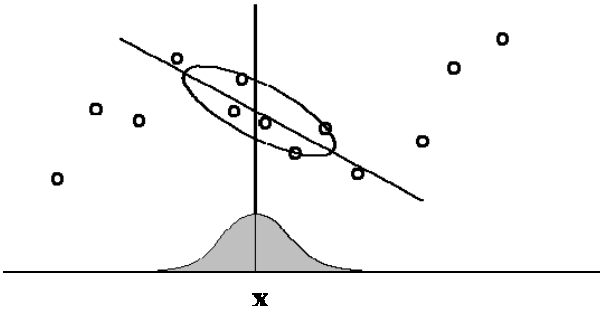

Figure 1 Gaussian kernel function

kernel function which is the most widely used (Mohammad et al, 2013).

2.3 Preparation of training instances

2.3.1 Method validation

To establish both models, we need to get enough training instances. The instances (total flavonoids concentration) were determined by ultraviolet spectrophotometry. We established the method to determine the concentration of total flavonoids in *Aurantii Fructus* fried with bran during the extracting process. Naringin was chosen as the standard substance, because its full wavelength of scan spectra during 200−800 nm was similar to the extracts of *Aurantii Fructus*, and the maximum absorption was 283 nm.

The methodological study (precision test, stability test, repeatability test, and recovery test) on the established determination method met the requirements. The standard curve and lower limit of quantification are as following. The linear relationship of absorbance (Y) and naringin mass concentration (X) is perfect, and the regression equation is $Y = 29.03\ X - 0.018\ 9$, $R^2 = 0.999\ 7$, and the lower limit of quantification is 0.002 87 − 0.034 44 g/L.

2.3.2 Preparation of samples

In our study, the total flavonoids concentration depends on the t and M. We need to fix one parameter to see the relation of C_B with the others. M = 10, 14, 20 were chosen to develop the function between t and C_B. We determined C_B under the different t of three solvent volumes. Then the effect of M was conducted and t = 30 and 100 min were chosen as the fixed factor. C_B was determined under the different M of two kinds of t.

At last we collected 38 samples to be the training datasets of the model. The data are shown in Table 2.

2.4 Mathematical modeling

2.4.1 Functional relation of t to C_B

Based on the training dataset and functional relation deduced in Table 1, M = 10, 14, 20 were chosen to develop the function between t and C_B. We determined the C_B under the different t of three solvent volumes. In the off-line analysis, it was found that the relation of natural logarithm of the total flavonoid concentration to natural logarithm of t was approximately linear, corresponding to the functional relation in formula 1-2. The plot can be drawn and the functional relations are shown in Figure 2.

Table 2 Training dataset used in modeling

t / min	M / (mL·g^{-1})	C_B / (mg·mL^{-1})	t / min	M / (mL·g^{-1})	C_B / (mg·mL^{-1})
10	10	3.2888	10	14	2.8238
20	10	4.9063	20	14	4.2441
30	10	6.0851	30	14	4.9802
40	10	6.8719	40	14	5.5638
60	10	7.9554	50	14	6.1201
80	10	8.5157	60	14	6.5979
100	10	9.4169	80	14	7.1877
120	10	10.0459	100	14	7.6580
135	10	10.2589	120	14	7.7656
150	10	10.4333	140	14	7.8871
165	10	10.6762	160	14	8.0777
180	10	10.5063	120	20	5.4931
10	20	1.8939	140	20	6.1848
20	20	2.7878	30	12	5.3839
30	20	3.3961	30	16	3.7927
40	20	3.8293	30	18	3.6644
60	20	4.4537	100	12	8.2457
80	20	5.0376	100	16	6.1553
100	20	5.2157	100	18	5.9054

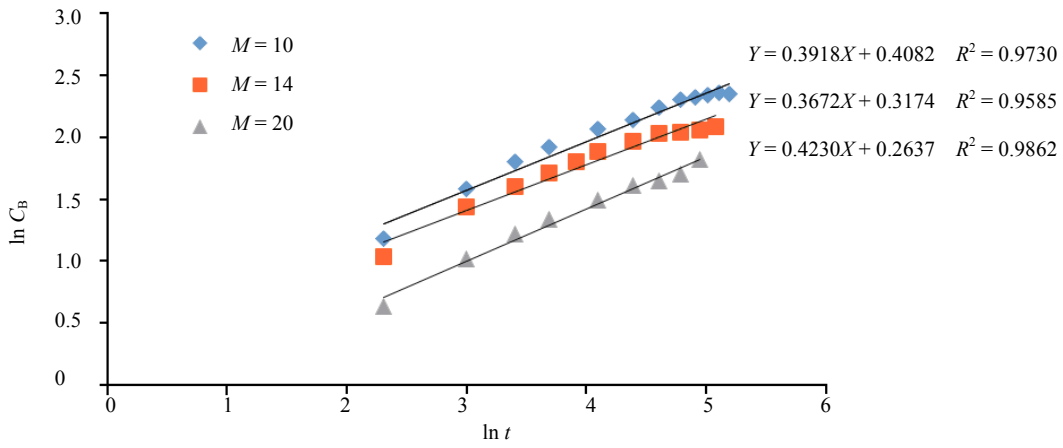

Figure 2 Functional relation of t to C_B

2.4.2 Functional relation of M to C_B

Then the effect of M was conducted and t = 30 and 100 min were chosen as the fixed factors. C_B was determined under different M of two kinds of t. It was found that the relation of natural logarithm of the total flavonoids concentration to the natural logarithm of M-R was approximately linear, which was corresponding to the functional relation in Table 1 (formula 1-3). After we got the water absorption coefficient R, the plot could be drawn, and the functional relations are shown in Figure 3.

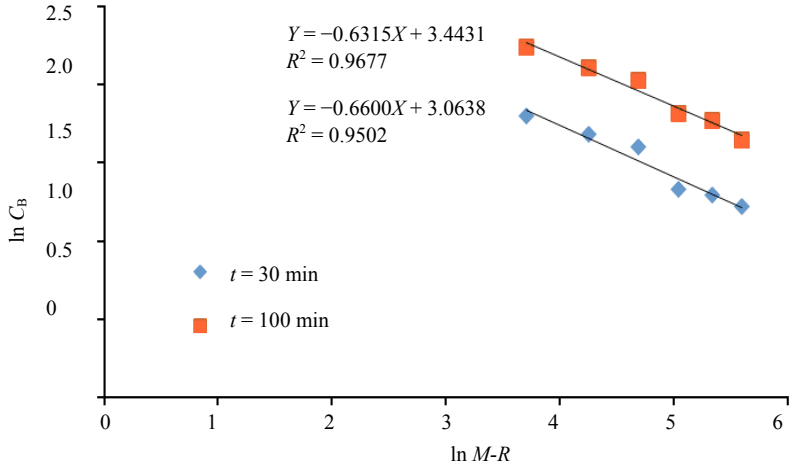

Figure 3 Functional relation of *M*-*R* to C_B

2.4.3 Establishment of equation

The linear equations in Figures 2 and 3 are corresponding to the equation deduced in Table 1 (formula 1-3 and 1-4). The parameters R and f_1 are fixed and their values can be measured.

In $f1 = \frac{2(C_{1b} - C_{10})}{\sqrt{\pi}}$, $(C_{1b} - C_{10})$ stands for the total flavonoids concentration of the medicinal herbs; R is water absorption coefficients, which means how much the volume of solvent per gram of medicinal herbs absorbed. So we got R = 3.626 mL/g, f_1 = 0.2424.

From Figure 2 we knew that M = 10, functional relation of t to C_B is Y = 0.3918X + 0.4082; M = 14, Y = 0.3672X + 0.3174; M = 20, Y = 0.4230X − 0.2637. These three functions were corresponding to the following equations.

$$\ln C_B = \frac{1}{1-n}\ln\frac{af_1}{M-R} + \frac{1}{2(1-n)}\ln t \qquad (1\text{-}3)$$

where $\frac{1}{1-n}\ln\frac{af_1}{M-R}$ as the intercept, and $\frac{1}{2(1-n)}$ as the slope.

So we could get n and a from each equation. Calculating the average of three n and a, we got n_1= −0.2733, a_1= 53.22.

Likewise, we could get another n and a from the functional relation of M to C_B, corresponding the equations in Figure 3 to the following equation.

$$\ln C_B = \frac{1}{1-n}\ln af_1 t^{\frac{1}{2}} - \frac{1}{(1-n)}\ln(M-R) \qquad (1\text{-}4)$$

where n_2 = −0.5492, and a_2 = 78.85, $n = (n_1 + n_2)/2$ = −0.3837, and $a = (a_1 + a_2)/2$ = 63.47.

If they were substituted into formula 1-2, we could obtain the kinetic model equation of total flavonoids extraction process of *Aurantii Fructus* as $C_B = (15.39 \times t^{0.5} / M - 3.626)^{0.7227}$.

2.5 LWR modeling

Training and testing of the LWR were carried out within Weka 3.6.10 using the same training subsets as mathematical modeling, and all the data were saved in a CSV document.

LWR uses an instance-based algorithm to assign instance weights which are then used by a specified weighted instances handler. And this can do the regression we need in the modeling.

The name of the learning is Weka.classifiers.lazy.LWL. In order to determine the optimal correlation coefficient (r), some parameters need to be set. (1) KNN, which means how many neighbors are used to determine the width of the weighting function; (2) Classifier, the base classifier to be used; (3) Nearest Neighbor Search Algorithm, the nearest neighbor search algorithm to use (Default: Linear NN); (4) Weighting Kernel, which determines weighting function (0 = linear, 1 = epnechnikov, 2 = tricube, 3 = inverse, 4 = gaussian, and 5 = constant).

Optimization of the parameters mentioned above is the most important work in this modeling.

The most important task is the value of KNN. We tried from 2−15 to get the optimal r value. The results strongly indicated that KNN = 5 could predict with the best accuracy. And the parameter 3 was set to Default: Linear NN; parameter 4 was set to 0 = linear in order to get the biggest correlation coefficient. So an LWR modeling was established at last. And the correlation coefficient is 0.993.

3. Results

After the establishment of the both models, we can

compare their accuracies in the prediction. Thirty testing subsets with different t and M were submitted and total flavonoids C_B of each combination was determined as the exact experimental data.

When we compared the prediction results of the both models with the exact experimental data, the performance was different. The statistical analysis results are shown in Table 3, including average percent relative error (E_r), average relative percent standard deviation (RSD), and r.

We use the cross plot to describe the performance with a reference line with a slope of 45° to ascertain the fitness and accuracy of the correlation. The error of each prediction is relative to the distance between each point and the diagonal line (Mohammad et al, 2013). A perfect correlation would show a straight line with a slope of 45° (Figure 4).

Table 3 Statistical accuracy of both models in predicting concentration of total flavonoids

Models	E_r	RSD / %	r
LWR modeling	−0.024	2.230	0.993
mathematical modeling	2.332	4.299	0.967 (average)

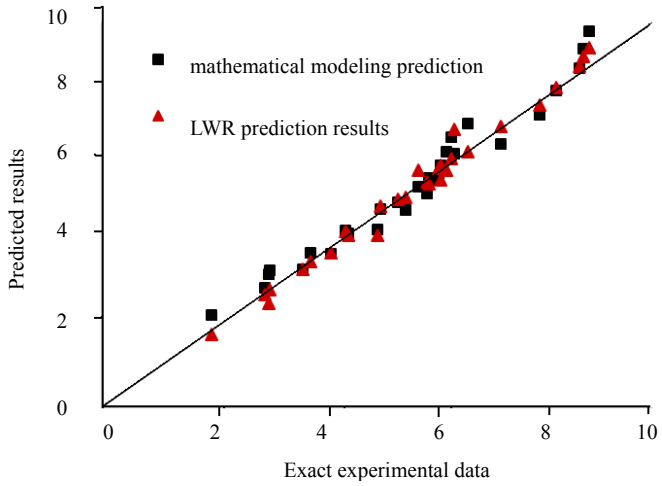

Figure 4 Comparison of exact experimental data with predicted value of both models for concentration of total flavonoids

4. Discussion

Table 3 and Figure 4 show that the both models did a good job in prediction. The RSD value was controlled into error range of Chinese pharmaceutical industry. They could be used in predicting the total flavonoids concentration in the extraction process of *Aurantii Fructus*. Under comparison, the LWR model using Weka has smaller error and bigger correlation coefficient. Its scatted points in figure were nearer to the diagonal line which means bigger overall accuracy.

In the mathematical modeling, two fixed parameters R and f_1 should be determined, but not need to be determined in LWR modeling. This means less work but more accuracy. However, the both models are deviated most from the diagonal line at the end of line. This is because that with the increase t, the concentration of total flavonoids from *Aurantii Fructus* is the same as the concentration in the extraction solvent; The extracting process is too slow to continue, So the mathematical equation could not fit this. As to LWR, no such things happened, just because there were not enough training sets around them. Only if we do more experiments to increase t, we could get a better accuracy.

5. Conclusion

This is the first time of establishing the kinetic model equation of total flavonoids extraction process of *Aurantii Fructus*: $C_B = (15.39 \times t^{0.5} / M - 3.626)^{0.7227}$, and also the first time of applying LWR theory into extraction process of traditional Chinese medicine. The results are promising and can be used to do the prediction in the extraction process of *Aurantii Fructus*.

For the mathematical model based on mass transfer theory and Fick's law, the equation could not fit when the concentration gradient no longer existed. Additionally, we need to deduce the mathematical relation at first, which needs mathematical and physical knowledge. However, the LWR model only needs enough training subsets. Of course, the experimental design is also important, because it helps to get the most effective training subsets. So, many problems can be solved for the researchers in other fields, who know little mathematical and physical knowledge. This is the amazing part of data mining. Also, it can be used into process prediction when the process has no clear rules or quantity relations to follow.

In this LWR modeling, the optimal results of the parameter 3 (Nearest Neighbor Search Algorithm) are Linear NN, which means we have choosen the locally weighted linear regression models this time. This may be related to the linear relationship deduced by mathematical model. But it is just one algorithm of LWR. Many other new algorithms can be explored in the other fields.

Reference

Bezerra MA, Santelli RE, 2008. Response surface methodology

(RSM) as a tool for optimization in analytical chemistry. *Talanta* 76: 965.

Chen F, Xing XF, Bai X, Liu SX, Zhang TJ, Chen CQ, 2014. Optimation of extraction technology for Xinning Effervescent Tablet by central composite design/response surface method. *Chin Tradit Herb Drugs* 45(11): 1566-1572.

Cost S, Salzberg SA, 1993. Weighted nearest neighbor algorithm for learning with symbolic features. *Machine Learning* 10(1): 57-78.

Firlbeck D, Faulstich M, Urmann C, Azaizeh H, Tafesh A, Riepl H, 2013. Central composite design for optimal technology of concentrating vanillic acid using foam fractionation. *Sep Purif Technol* 119: 28.

Eibe F, Mark H, Bernhard P, 2003. Locally Weighted Naive Bayes. *In: 19th Conference in Uncertainty in Artificial Intelligence*, Acapulco.

Han LX, Ni J, Xuan L, Dou JJ, 2011. Kinetic model for extraction of *Acanthopanax Senticosus*. *Chin Traditi Patent Med* 33(2): 240-245.

Hou KF, Zheng Q, Li YR, Shen JZ, Hu SY, 2000. Modeling and optimization of herb leaching processes. *Comput Chem Eng* 24(2/7): 1343-1348.

Ian HW, 2006. *Data mining*. 1st ed. China Machine Press, Beijing.

Li YR, Cheng S, 1997. Simulation and optimization of herb extraction process. *Chin Tradit Herb Drugs* 28(7): 399-401.

Mai LY, Li YX, Chen Y, Li J, Zhong MY, 2014. Applications of mathematical statistics methods on compatibility researches of traditional Chinese medicines formulae. *China J Chin Mater Med* 39(10): 1749-1756.

Mei CL, Zhang WX, 2001. Statistical inferrences for varying-coefficint models based on locally weighted regression technique. *Acta Math Appl Sin* 17(3): 98-101.

Mohammad SL, Majid T, Bahman M, 2013. Viscosity prediction in selected Iranian light oil reservoirs: Artificial neural network versus empirical correlation. *Petroleum Sci* 10: 126-133.

Siepmann J, Siepmann F, 2008. Mathematical modeling of drug delivery. *Int J Pharm* 364: 328-343.

Spiro M, Selwood RM, 1982. The kinetics and equilibria of tea infusion. Part 3. Rotating disc experiments interpreted by a steady state mode . *J Chem Soc Faraday Trans* 78: 295-305.

Su H, Wang BC, Liu WQ, Qian JZ, Yang X, Li W, 2011. Kinetic mathematical models for extraction process of natural medicines. *Chin Tradit Herb Drugs* 42(2): 384-391.

WangYK, Liu XG, 2012. Modeling hot metal silicon content in blast furnace based on locally weighted SVR and mutual information. *Thirty-first Chinese Control Conferenc*e 7089-7094.

William SC, Susan JD, 1988. Locally weighted regression: An approach to regression analysis by local fitting. *JASA* 403(83): 596-610.

Wu CL, Liu Y, Zhang ZF, Xu Y, 2011. Optimization of extracting condition for Morina nepalensis by uniform design. *Chin Tradit Herb Drugs* 42(12): 2447-2450.

Zhang J, Chen Y, Cao SL, Yin XB, Qi JJ, Zhang H, Dang XF, Ni J, 2013a. Study on applicability of kinetic model for water extracts from *Puerariae Radix*. *China J Chin Mater Med* 38(17): 2788-2792.

Zhang WL, Zhang SF, Fan HL, Yu HY, Jing WL, 2013b. Orthogonal experimental design for the optimal ingredient composition of modified Zexietang in the treatment of hypertension. *World J Integr Tradit West Med* 8(2): 128-131.