Pharmacogn. Mag.
A multifaceted peer reviewed journal in the field of Pharmacognosy and Natural Products
www.phcog.com | www.phcog.net

ORIGINAL ARTICLE

Rapid and Simultaneous Analysis of Five Free Anthraquinone Contents in Rhubarb during the Stir-Frying with Rice Wine Process by Near Infrared Reflectance Spectroscopy

Xingyi Chen[1], Jintao Xue[2], Weiying Wang[1], Bihua Ma[1], Guo Huang[3], Yu Chen[4], Liming Ye[1]

[1]Department of Pharmaceutical Analysis, West China School of Pharmacy, Sichuan University, [3]Department of Application, Sichuan Vspec Technologies Co., Ltd., [4]Department of Pharmacy, West China Hospital, Sichuan University, Sichuan, [2]Department of TCM, School of Pharmacy, Xinxiang Medical University, Henan, China

Submitted: 10-11-2018 Revised: 17-12-2018 Published: 16-05-2019

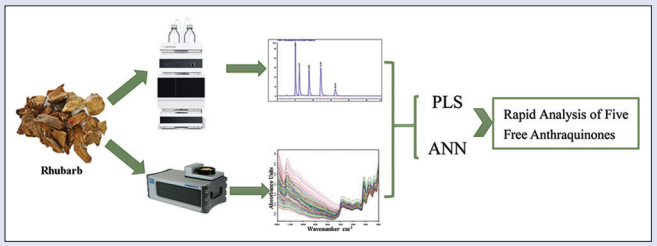
ABSTRACT

Background: The active ingredients of Traditional Chinese Medicines (TCMs) vary greatly with the degree of stir-frying; so, rapid analysis of the active content is very important for the processing of TCMs. **Objective**: In this study, near infrared reflectance (NIR) spectroscopy was used to develop a new method for the rapid online analysis of five free anthraquinones (aloe-emodin, rhein, emodin, chrysophanol, and physcion) during the stir-frying process for rhubarb. **Materials and Methods:** With partial least-squares (PLSs) and artificial neural networks (ANN) regression, calibration models were generated based on five free anthraquinone contents, as measured by high-performance liquid chromatography. **Results:** The results indicated that the 2 types of models were robust, accurate, and repeatable for five free anthraquinones. Moreover, PLS as the linear model was more suitable for developing the NIR models of the five free anthraquinones than ANN. The performance of the optimal models was achieved as follows: the coefficient of determination for prediction (R^2_{pre}) for aloe-emodin, rhein, emodin, chrysophanol, and physcion was 0.9161, 0.9699, 0.9655, 0.9611, and 0.9724, respectively; the root mean square error of prediction was 0.0251, 0.0445, 0.3333, 0.0862, and 0.0211, respectively. **Conclusion:** The established NIR models could apply to determine the content of five free anthraquinones in rhubarb. This work demonstrated that NIR may be an effective online analysis method to reflect the quality of TCM industrial manufacturing processes.

Key words: Artificial neural network, anthraquinones, near infrared reflectance spectroscopy, partial least squares, rhubarb

SUMMARY

- Near infrared reflectance (NIR) spectroscopy was used to develop a method for rapid online analysis of free anthraquinones in rhubarb
- With partial least-square and artificial neural networkregression, calibration models of NIR were generated based on five free anthraquinone contents measured by high-performance liquid chromatography
- The 2 calibration models were robust, accurate, and repeatable. NIR may be an effective online analysis method to reflect the quality of the Traditional Chinese Medicine industrial manufacturing process.



Abbreviations used: ANN: Artificial neural networks; COE: Constant offset elimination; D: Dimension; FDA: U. S. Food and Drug Administration; FT-NIR: Fourier transform near infrared reflectance; HPLC: High-performance liquid chromatography; LC-MS: Liquid chromatography-mass spectrometry; MLPs: Multilayer perceptrons; MMN: Min/Max normalization; MSC: Multiplicative scatter correction; MSE: Mean square error; NIR: Near infrared reflectance; OE: Offset elimination; PAT: Process analytical technology; PCA: Principal components analysis; PLS: Partial least squares; R^2cal: Correlation coefficient of the calibration set; R^2pre: Coefficient of determination for prediction; RMSECV: Root mean square error of cross-validation; RMSEP: The root mean square errors of prediction; SD: Standards deviation; SLS: Straight line subtraction; SNV: Standard normal variate transformation; TCM: Traditional Chinese medicine; 1st D: First derivative; 2nd D: Second derivative.

Correspondence:

Prof. Liming Ye,
West China School of Pharmacy,
Sichuan University, No. 17 People's South Road,
Chengdu 610041, Sichuan, P. R. China.
E-mail: yeliminglaoshi@126.com
DOI: 10.4103/pm.pm_562_18

Access this article online
Website: www.phcog.com
Quick Response Code:

INTRODUCTION

According to the ICH Q8, it is not recommended that the quality of drugs is assured by testing, but by design and production.[1] The US Food and Drug Administration also advocated the use of process analytical technology (PAT) to improve pharmaceutical manufacturing and assure the quality of pharmaceutical production, which provided a reference for Traditional Chinese Medicine (TCM) production.[2,3] In TCM production, one of the most urgent problems is uniform product quality. The current method for identifying the degree of processing is mostly based on appearance, which is not only subjective but also cannot reflect the content changes accurately.[4] One of the reasons TCMs has not gained acceptance in overseas markets is the inconsistent quality of TCM.

As an attractive PAT tool, near infrared reflectance (NIR) spectroscopy has been widely used in chemical and biological drug production.[3] In recent years, the technology has also been introduced into TCM industries,

Cite this article as: Chen X, Xue J, Wang W, Ma B, Huang G, Ye L, *et al*. Rapid and simultaneous analysis of five free anthraquinone contents in rhubarb during the stir-frying with rice wine process by near infrared reflectance spectroscopy. Phcog Mag 2019;15:479-86.

including raw materials[5-7] and process monitoring.[8-10] In process monitoring, the most common methods of quantitative analysis are high-performance liquid chromatography (HPLC) and liquid chromatography mass spectrometry (LC-MS). However, online analysis is not ideal for controlling TCM processing production, as it requires tedious sample preparation, time-consuming sample analysis, and large amounts of solvent. NIR is a rapid analytical technique that can produce a spectrum in a few seconds. The most significant advantage of this technique is the nondestructive character of the analysis: a sample can be analyzed without or with only little sample preparation, maintaining the integrity of samples and saving preprocessing time.[11] It is even possible to measure packaged samples through the package material. Moreover, compared to HPLC and LC-MS, NIR spectra provide more physical and chemical information. The above advantages show that NIR is an ideal PAT tool.

NIR absorbance peaks were relatively weak and highly overlapping; so, extracting effective information from NIR spectra and optimizing the raw spectrum data are essential for establishing an ideal model in NIR. Many multivariate data analysis methods have been used for developing NIR mathematical models, such as partial least-squares (PLS), artificial neural networks (ANN), multiple linear regression, and principal component analysis.[12-15] Among these methods, PLS and ANN were the main study methods.

Rhubarb is a component in TCMs that are widely used for cathartic, antidotal, and febrifugal purposes.[16,17] Rhubarb, a species from *Polygonacea* in the genus *Rheum*, was identified as the dried rhizome and root of *Rheum plamatum* L., *Rheum tanguticum* Maxim. ex Balf., and *Rheum officinale* Baill. in Chinese pharmacopoeia.[18] In general, rhubarb must be processed to alleviate its strong laxative effect. At present, stir-frying with rice wine is the most commonly used method in rhubarb processing, in which rhubarb is saturated with rice wine for several hours and stir-fried to dryness.[19,20]

At present, the application of NIR in rhubarb focuses on qualitative analysis that could distinguish official and unofficial samples.[21-23] However, they cannot reflect the content of active ingredients in rhubarb concretely. This research mainly reports quantitative analysis with an NIR model in rhubarb. During the stir-frying of rhubarb, the content of free and combined anthraquinones is changed. In this study, the five free anthraquinones (aloe-emodin, rhein, emodin, chrysophanol, and physcion) were chosen as the detection index for stir-frying process. As NIR had shown enormous potential and gained wide acceptance for the analysis of TCM, it is urgent to establish NIR method for the quality analysis of rhubarb.

In this study, six batches of rhubarb collected from different origins were used to develop the NIR models. Two types of modeling were applied: PLS as linear regression and ANN as nonlinear regression. The results indicated that both of them were robust, accurate, and repeatable for online analysis and quality control. Moreover, PLS performed better in the development of NIR models of five free anthraquinones than did ANN. It showed that NIR has great potential in the TCM industrial manufacturing process.

MATERIALS AND METHODS

Materials and reagents

Six types of rhubarb were obtained from 4 provinces in China (Sichuan, Gansu, Shaanxi, and Qinghai). Aloe-emodin (purity >98.1% by HPLC), rhein (purity >99.3% by HPLC), emodin (purity >98.7% by HPLC), chrysophanol (purity >99.2% by HPLC), and physcion (purity >99.0% by HPLC) were from the National Institute for Food and Drug Control (Beijing, PR China). Rice wine was purchased from the Kuaijishan Shaoxing Wine Company (Shaoxing, Zhejiang, PR China). HPLC-grade methanol was obtained from Tianjin Kermel Chemical Reagent Company (Tianjing, PR China). Water was purified by an ultrapure water instrument. All other reagents were of analytical grade.

Sample preparation

One type of rhubarb was randomly selected from six types to research the main factors affecting the process and the optimum process condition. According to the *Chinese Pharmacopoeia* (volume II, 2015 edition), stir-frying temperature, rice wine content, and sealed moistening time were chosen as the main research factors for the optimum process conditions. Using the content of total free anthraquinones as a reference, the optimum process conditions were determined using the orthogonal test design of L_{25} (5^3) [Table 1]. According to orthogonal test results, the stir-frying temperature, as the single control factor, was used to research the change in rhubarb during the process. Each batch of rhubarb received 7.5% (v/v) rice wine and was sealed for 3 h, and the stir-frying samples were collected at 30°C, 90°C, 110°C, 130°C, 150°C, 170°C, 180°C, 190°C, 200°C, and 210°C, respectively. The temperature of the samples was controlled using a GM320 infrared thermometer (BENETECH Company, Shenzhen, Guangdong, RP China). Therefore, each type of rhubarb contained 3 batches, and each batch contained 10 degrees of stir-frying samples, except that the type of rhubarb used for the orthogonal test had only 2 batches.

All samples were milled into 100-mesh powder and dried in a silica gel desiccator for at least 7 h at room temperature (30°C) until the weight loss was <0.0003 g to ensure that moisture was not an interfering factor.

Near infrared reflectance spectrum collection

NIR spectra were acquired using a QuasIR 3000™ FT-NIR spectrometer (Galaxy Scientific Inc., Nashua, New Hampshire, USA) equipped with a 98mm sample cup and a sample spinner. The NIR spectrometer was operated using eFTIR software (Essential FTIR V3.00.047). The spectra were acquired at a resolution of 8 cm^{-1} over a wavelength range of 12000–4000 cm^{-1} with 32 scans per spectrum. The samples were scanned as powders. The dataset consisted of 170 samples, and two spectra were registered per sample [Figure 1a]. The averaged spectrum was used for computations.

High-performance liquid chromatography reference value collection

According to the Chinese Pharmacopoeia (volume II, 2015 edition), the Agilent 1100 HPLC system (Agilent Technologies Inc., USA) consisting of a UV–Vis detector was used for the quantitative determination of five free anthraquinones. A stir-frying sample was extracted for 1 h with 25 mL of methanol in a round-bottomed flask and filtered, and the subsequent solution was used for determination. Free anthraquinone determination was performed under isocratic conditions, with a mobile phase of methanol – 0.1% phosphorous acid (85:15) at a flow rate of 1 mL/

Table 1: Orthogonal design of processed rhubarb

Levels	Factors		
	A: Contents of rice wine (%)	B: Sealed moistening time (h)	C: Stir-frying temperature (°C)
1	5	1	90
2	7.5	3	120
3	10	6	150
4	12.5	9	180
5	15	12	210

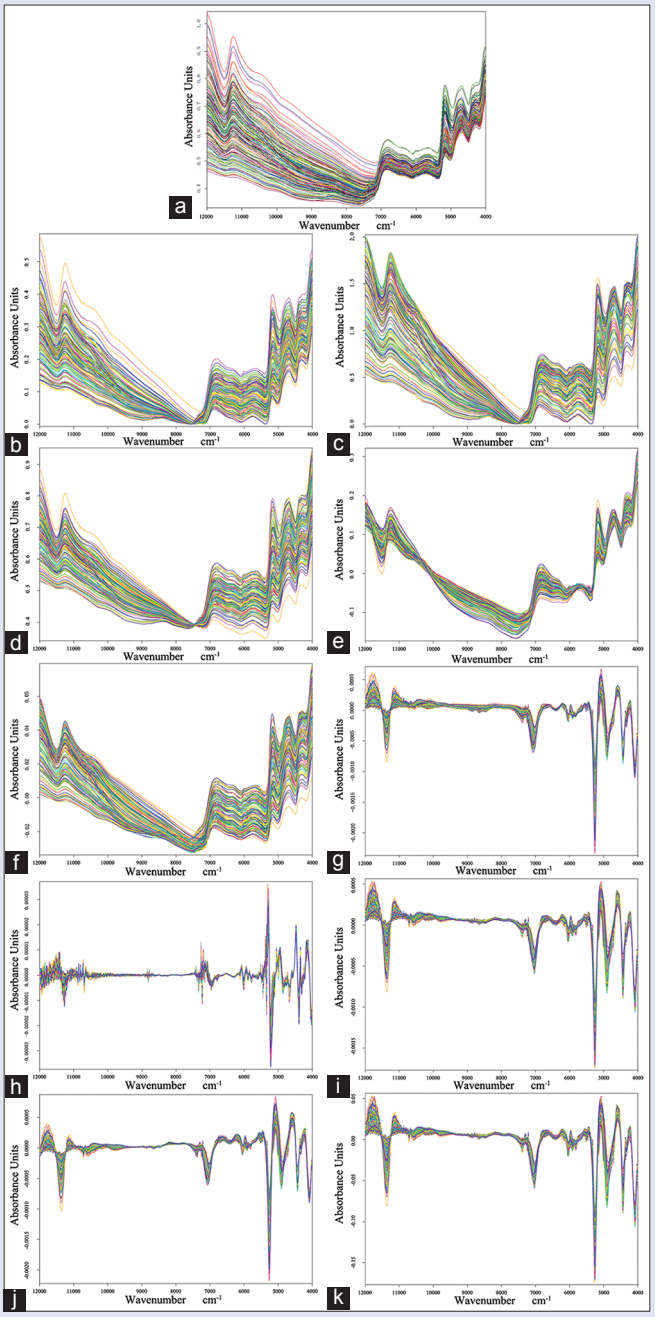


Figure 1: Spectra of different spectral pre-processing methods for PLS and ANN. (a) Raw specta; (b) constant offset elimination; (c) Min/max normalization; (d) MSC; (e) Straight line subtraction; (f) Standard normal variate transformation; (g) 1st D; (h) 2nd D; (i) 1st D + MSC; (j) 1st D + SLS; (k) 1st D + SNV. 1st D: Fist derivative; 2nd D: Second derivative; MSC: Multiplicative scatter correction; SLS: Straight line subtraction; SNV: Standard normal variate transformation; PLS: Partial least square; ANN: Artificial neural network

min at 35°C. Separation was performed on a Kromasil C_{18} column (5 μm, 250 mm × 4.6 mm) and monitored by UV detection (UVD; λ = 254 nm). The injection volume was 10 μL. The identification of five free anthraquinones (aloe-emodin, rhein, emodin, chrysophanol, and physcion) was based on retention time, and quantification was achieved by external standard calibration.

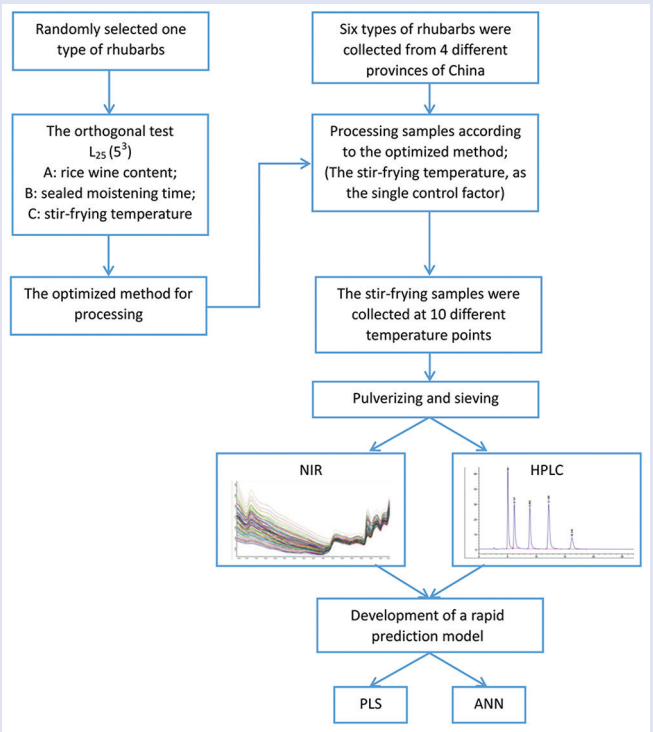


Figure 2: Specific operating process and methods in this study

Spectral preprocessing

The development of NIR calibration models primarily consists of two types: linear regression which heavily relies on a linear relationship between the reference value and the intensity of the NIR spectra and PLS, which is the most widely used method in this type; another is nonlinear regression and ANN was most widely used to reflect the nonlinear effects.[24-27] In this study, the linear relationship was established by PLS models with OPUS 7.5 software (Bruker Optik, Ettlingen, Germany). The nonlinear relationship was established by ANN models with NeuroSolutions 7 (Neurodimension Inc., Gainesville, USA). The specific operating process and methods are introduced briefly in Figure 2.

To obtain the best NIR models, different spectral pretreatments were tested to reduce unwanted variation due to sources not related to the properties of interest. The spectral pretreatment methods of PLS contain 10 methods including constant offset elimination (COE), min/max normalization (MMN), multiplicative scatter correction (MSC), straight line subtraction (SLS), standard normal variate transformation (SNV), fist derivative (1st D), second derivative (2nd D), 1st D + MSC, 1st D + SLS, and 1st D + SNV. These preprocessing methods were investigated to establish the model for five free anthraquinones.

The most widely used spectral preprocessing methods for ANN model were OE, MMN, and SNV. Moreover, all spectra were smoothed using the Savitzky–Golay algorithm first to improve spectral smoothness and reduce the interference of noise. The optimum smooth points were 13.

RESULTS AND DISCUSSION

Optimized method of processing

In accordance with the L_{25} (5^3) orthogonal experiment designed for the three factors (stir-frying temperature, rice wine content, and sealed moistening time) and five levels, the optimal process condition was

obtained using the content of total free anthraquinones as a reference. As shown in Table 2, the primary factor and minor factor influencing the content of five free anthraquinones during the stir-frying process were determined by range analysis. *R* was the level range that reflected the amplitude of the fluctuation test index when the corresponding factor was changed. A larger value of *R* indicates a greater influence of this factor. The result indicated that the stir-frying temperature was the primary factor, the rice wine content was the second factor, and the sealed moistening time was the last factor. *K* was the sum of the measuring values at the same level for each factor. A larger value of *K* indicates a better level. This indicated that the optimized method of processing was $A_2B_2C_5$. In other words, the optimized methods were adding 7.5% (v/v) rice wine, sealing for 3 h and stir-frying at 210°C.

High-performance liquid chromatography reference data

The five free anthraquinones of rhubarb at different degrees of stir-frying vary greatly. As shown in Table 3, with increasing temperature, the overall trend in the content of five free anthraquinones increased. Compared with the contents of five free anthraquinones from unprocessed samples, there were significant differences between the unprocessed and processed groups ($P < 0.05$). The change in chrysophanol was most significant in the five free anthraquinone, increasing approximately 0.2% during the stir-frying process.

Table 2: Results of L_{25} (5^3) orthogonal test

K	Factors		
	A	B	C
K_1	6.68	6.59	2.48
K_2	7.43	7.38	2.30
K_3	5.76	6.66	4.45
K_4	6.38	6.21	11.41
K_5	7.14	6.55	12.75
R	0.34	0.24	2.09
Optimal plan	A_2	B_2	C_5

Development of the partial least-square model

In PLS model, all 170 samples were randomly divided into two subsets. The first subset was called calibration set with 136 samples to be used to establish the calibration model, while the other one was called prediction set with 34 samples to be used for testing the robustness of model. Table 4 shows the HPLC reference data in calibration and prediction sets. As seen in Table 4, the range of reference data of each anthraquinone in the calibration set almost covers the range in the prediction set and their standards deviation (SD) between the calibration and prediction sets are no significant differences. Therefore, the distribution of the samples is appropriate both in the calibration and the prediction sets.

The performance of the PLS model was evaluated according to four types of parameters, i. e., the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP), and the coefficient of determination for calibration and prediction (R^2_{cal}, R^2_{pre}). The optimal model was selected based on the higher R^2_{cal}, R^2_{pre}, as well as a lower RMSECV and RMSEP.[15] Table 5 shows the result of the PLS calibration model, and the optimized models of five free anthraquinones are highlighted in bold.

The three major parameters for developing the PLS model are the spectral preprocessing method, the spectral region, and the dimension (D) factor. The spectral pretreatment methods of PLS contain 10 methods in total, including COE, MMN, MSC, SLS, SNV, 1st D, 2nd D, 1st D + MSC, 1st D + SLS, and 1st D + SNV. All were applied to each NIR spectrum to shift the baseline, enhance the spectral features, and eliminate noise and matrix background interference so that the relevant information can be extracted more fully. These preprocessing methods were applied to establish the model for five free anthraquinones. Figure 1 shows that spectra were pretreated with these methods. The optimal preprocessing methods for aloe-emodin, emodin, rhein, physcion, and chrysophanol were 1st D + MSC, 2nd, 1st + MSC, SLS, and 1st + SLS, respectively.

The different spectral regions contain different spectral signals. Extracting the effective wavelength range can improve the efficiency and accuracy of calibration models. To select the optimal NIR spectral region for the PLS model, at first, several wavelength ranges were extracted

Table 3: Content changes of five free anthraquinones between unprocessed and stir-frying rhubarbs

Content (%)	Unprocessed	Stir-frying temperature (°C)									
		30	90	110	130	150	170	180	190	200	210
Aloe-emodin	0.1536[a]	0.1496	0.1608	0.1657	0.1718	0.1725	0.1881	0.1851	0.1993	0.2154	0.2059
Rhein	0.1947[b]	0.1661	0.1845	0.1951	0.1958	0.2081	0.2111	0.2273	0.2272	0.2645	0.2507
Emodin	0.1857[c]	0.1807	0.1898	0.1902	0.2007	0.2006	0.2173	0.2238	0.2387	0.2536	0.3099
Chrysophanol	0.6878[d]	0.7215	0.7375	0.7811	0.8248	0.7871	0.8521	0.8299	0.9321	0.9509	0.9442
Physcion	0.2383[f]	0.2011	0.2081	0.2181	0.2291	0.2192	0.2494	0.2449	0.2762	0.2933	0.2811

[a]$P<0.05$; [b]$P<0.05$; [c]$P<0.05$; [d]$P<0.05$; [f]$P<0.05$, compared with processed group

Table 4: References data in the calibration and the prediction set

Subsets	Components (%)	PLS			ANN		
		Range	Mean	SD	Range	Mean	SD
Calibration set	Aloe-emodin	0.0430-0.3197	0.1854	0.0569	0.0430-0.3197	0.1854	0.0576
	Rhein	0.0079-0.6066	0.2099	0.1935	0.0079-0.6066	0.2035	0.1898
	Emodin	0.0360-0.5430	0.2250	0.1240	0.0368-0.5391	0.2184	0.1232
	Chrysophanol	0.1013-1.7257	0.8823	0.2990	0.1082-1.7257	0.8753	0.2994
	Physcion	0.0479-0.5766	0.2454	0.0991	0.0479-0.5766	0.2405	0.0976
Prediction set	Aloe-emodin	0.0516-0.3001	0.1800	0.0623	0.0452-0.2537	0.1755	0.0609
	Rhein	0.0105-0.5191	0.2010	0.1849	0.0109-0.5363	0.2468	0.2050
	Emodin	0.0453-0.5091	0.2098	0.1266	0.0360-0.5430	0.2537	0.1366
	Chrysophanol	0.1281-1.4048	0.8269	0.3060	0.1013-1.2718	0.8372	0.3148
	Physcion	0.0601-0.4945	0.2297	0.0918	0.0480-0.4295	0.2573	0.0999

SD: Standard deviation; PLS: Partial least squares; ANN: Artificial neural networks

Table 5: Results of different spectral pretreatment methods in partial least squares model

Contents	Pretreatment method	Spectral range (cm^{-1})	D	Calibration set		Prediction set	
				RMSECV	R^2_{cal}	RMSEP	R^2_{pre}
Aloe-emodin	1st D + MSC	7501-5446.2, 4602.2-4247	15	0.0205	0.8625	0.0251	0.9161
	MSC	9403-6099.3, 5450.1-4598.4	12	0.0200	0.8620	0.0390	0.8145
	1st D	9403-7497.2, 6103.1-4598.4	18	0.0210	0.8535	0.0398	0.8108
	1st D + SLS	9403-7497.2, 6103.1-5446.2, 4602.2-4247	15	0.0220	0.8414	0.0353	0.8319
	2nd	9403-4247	13	0.0217	0.8382	0.0350	0.8317
	1st + SNV	9403-5446.2, 4602.2-4247	18	0.0230	0.8310	0.0288	0.8977
	SLS	6103.1-4598.4	18	0.0227	0.8255	0.0294	0.8834
	SNV	7501-5446.2, 4602.2-4247	19	0.0236	0.8176	0.0281	0.8983
	COE	7501-5446.2	19	0.0237	0.8070	0.0342	0.8463
	MMN	7501-6099.3, 4602.2-4247	12	0.0242	0.8012	0.0364	0.8121
Rhein	2nd	7501-4247	16	0.0322	0.9708	0.0445	0.9699
	1st + SNV	7501-5446.2	18	0.0366	0.9620	0.0529	0.9570
	1st + SLS	8452-7497.2, 6103.1-4247	19	0.0366	0.9614	0.0426	0.9739
	COE	9403-7497.2, 6103.1-5446.2	16	0.0367	0.9614	0.0907	0.8706
	1st	6103.1-4247	19	0.0389	0.9572	0.0456	0.9693
	1st + MSC	7501-5446.2, 4602.2-4247	17	0.0390	0.9559	0.0468	0.9676
	SNV	6103.1-4247	19	0.0409	0.9522	0.0586	0.9497
	MSC	6103.1-4247	18	0.0415	0.9512	0.0560	0.9534
	SLS	9403-7497.2, 6103.1-4598.4	17	0.0418	0.9507	0.0534	0.9577
	MMN	6103.1-4598.4	20	0.0443	0.9435	0.0522	0.9591
Emodin	1st + MSC	9403-5446.2, 4602.2-4247	10	0.0232	0.9624	0.0333	0.9655
	MMN	9403-4598.4	16	0.0250	0.9569	0.0534	0.9255
	1st + SNV	9403-5446.2	13	0.0264	0.9550	0.0372	0.9564
	SNV	6103.1-4598.4	17	0.0273	0.9500	0.0295	0.9721
	MSC	6103.1-4598.4	16	0.0277	0.9481	0.0297	0.9719
	COE	6103.1-4598.4	17	0.0287	0.9450	0.0330	0.9652
	1st	9403-7497.2, 6103.1-4247	17	0.0303	0.9419	0.0508	0.9334
	SLS	9403-6099.3, 4602.2-4247	16	0.0306	0.9362	0.0473	0.9324
	2nd	6103.1-4247	14	0.0327	0.9301	0.0269	0.9788
	1st + SLS	9403-5446.2	14	0.0333	0.9284	0.0411	0.9464
Chrysophanol	SLS	9403-7497.2, 6103.1-5446.2, 4602.2-4247	12	0.0643	0.9527	0.0862	0.9611
	1st + MSC	7501-5446.2, 4602.2-4247	13	0.0603	0.9591	0.1170	0.9275
	MSC	7501-4247	15	0.0620	0.9572	0.1190	0.9237
	SNV	7501-5446.2, 4426.5-4247	16	0.0534	0.9684	0.1250	0.9137
	MMN	9403-7497.2, 6103.1-5446.2, 4602.2-4247	15	0.0663	0.9487	0.1060	0.9368
	1st + SNV	9403-5446.2	11	0.0681	0.9467	0.1100	0.9381
	1st + SLS	9403-5446.2, 4602.2-4247	11	0.0714	0.9414	0.1100	0.9374
	COE	6103.1-5446.2, 4602.2-4247	8	0.0760	0.9330	0.0919	0.9555
	2nd	7501-4247	13	0.0772	0.9297	0.0910	0.9584
	1st	9403-7497.2, 6103.1-5446.2, 4602.2-4247	11	0.0809	0.9226	0.0835	0.9624
Physcion	1st + SLS	9403-5446.2, 4602.2-4247	13	0.0199	0.9596	0.0211	0.9724
	MSC	9403-6099.3, 5450.1-4247	17	0.0193	0.9602	0.0441	0.8795
	1st + MSC	7501-5446.2, 4602.2-4247	14	0.0196	0.9600	0.0227	0.9681
	1st + SNV	9403-5446.2, 4602.2-4247	12	0.0197	0.9610	0.0245	0.9628
	MMN	9403-5446.2, 4602.2-4247	15	0.0196	0.9588	0.0701	0.7450
	1st	9403-5446.2, 4602.2-4247	14	0.0203	0.9583	0.0293	0.9475
	SNV	7501-5446.2	18	0.0210	0.9555	0.0247	0.9631
	SLS	9403-6099.3, 5450.1-4247	19	0.0202	0.9551	0.0285	0.9502
	2nd	7501-6099.3, 4602.2-4247	13	0.0209	0.9525	0.0286	0.9595
	COE	9403-5446.2, 4602.2-4247	17	0.0217	0.9525	0.0295	0.9465

OE: Offset elimination; COE: Constant OE; MMN: Min/max normalization; MSC: Multiplicative scatter correction; SLS: Straight line subtraction; SNV: Standard normal variate transformation; 1st D: Fist derivative; 2st D second derivative; RMSEP: Root mean square error of prediction; RMSECV: Root mean square error of cross-validation

from the basis of the peak valley as shown in Figure 1a and then were used to establish the model; ultimately, the region that had a better performance was chosen. As shown in Table 5, different spectral regions were examined for the five free anthraquinones. The best spectral regions for each anthraquinone were as follows: aloe-emodin - 7501–5446.2 and 4602.2-4247 cm^{-1}; rhein - 7501-4247 cm^{-1}; emodin - 9403–5446.2 and 4602.2–4247 cm^{-1}; chrysophanol - 9403–7497.2, 6103.1–5446.2 and 4602.2-4247 cm^{-1}, and physcion - 9403–5446.2 and 4602.2-4247 cm^{-1}.

Each PLS model has an optimized D. In other words, if D is too low or too high, it will cause "under-fitting"; if D is not sufficient to explain the complex variability of the spectral data, this is "over-fitting" and spurious incorporated noise will be introduced.[28-30] Therefore, either lower or greater D will decrease the predictability, robustness, and accuracy of the PLS model. Figure 3 shows changes in R^2_{cal} with D for the model of physcion. The value of R^2_{cal} increased with increased D, which indicates that the model became increasingly robust. The optimal D for aloe-emodin, rhein, emodin, chrysophanol, and physcion was 15, 16, 10, 12, and 13, respectively.

Development of the artificial neural network model

The ANN model data were also randomly divided into two subsets: the calibration set and prediction set. However, the calibration set contained two groups: the training subset and test subset. 151 samples were selected randomly as a calibration set; 128 served as a training subset used to estimate ANN model parameters; 23 served as the test subset used to test the calibration model during the training process; and the 19 remaining samples were taken as the prediction set. The HPLC reference data in calibration and prediction sets are shown in Table 4.

Similarly, the performance of ANN models was assessed by MSE, RMSEP, R^2_{cal}, and R^2_{pre}. The optimum models should have lower MSE and RMSE*P* values, as well as higher R^2_{cal} and R^2_{pre} values.[31,32] The ANN models of five free anthraquinones are listed in Table 6, and the optimized model is highlighted in bold.

The major factors of the ANN model are different from the PLS model, including the spectral preprocessing method, the number of hidden neuron, transfer function, and processing elements (step size, momentum and iterations). The most common spectral pretreatment methods of the ANN model were OE, MMN and SNV. Even using the same test method during the development of the ANN model, the intensity of spectra may vary depending on the amount of samples used. With preprocessing, the spectra can be compared more accurately. The optimal preprocessing methods for aloe-emodin, emodin, rhein, physcion, and chrysophanol were SNV, MMN, SNV, SNV, and SNV, respectively. Moreover, in order to extract effective information and improve the running speed of ANN model, 200 wavelength points were extracted from the basis of the peak valley. The selecting wavelength points were concentrated in the wavelength range of 7000-4000cm^{-1} which contained abundant information.

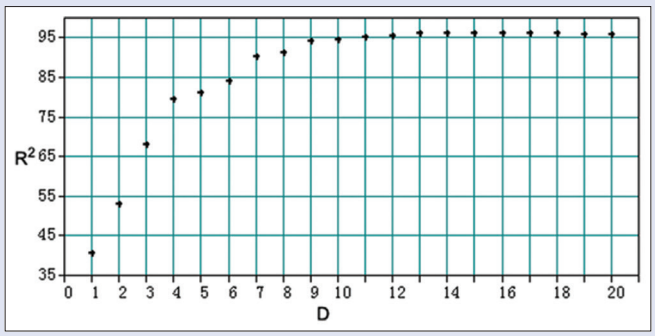


Figure 3: Relationship between R^2_{cal} and D of physcion in the partial least-square model

Another parameter to be optimized in the ANN model is the number of nodes in the hidden layer. In principle, the number of hidden neurons determines the complexity of the network. If the number of hidden units is insufficient, prediction errors tend to be elevated due to poor model fitting. However, too many hidden units can cause overfitting, increasing prediction errors. The method to determine the optimum number of hidden neuron is to start with a minimum number and continually add one new unit.[33] The results showed that the optimum number of hidden neurons for aloe-emodin, emodin, rhein, physcion, and chrysophanol was 8, 12, 12, 10, and 9, respectively.

The transfer functions most widely used in hidden layers are the sigmoid Axon or tanh Axon, which enable many nonlinearities to be fitted. The transfer functions of five free anthraquinones were all tanh Axon.

In this study, multilayer perceptrons, the most common network topology, was applied; these are layered feed forward networks typically trained with static backpropagation. Their main advantages are that they are easy to use and can be implemented efficiently with computers. Back propagation is slow to converge but can be improved by selecting the appropriate processing elements. For example, a larger step size allows the minimum to be reached more rapidly. However, if the step size is too large, the algorithm will diverge and the error will increase instead of decreasing. If the step size is too small, it will take too long to reach the minimum, which also increases the probability of getting caught in local minima. The optimum processing elements of the five free anthraquinones are shown in Table 6.

Table 6: Results of different spectral pretreatment methods in artificial neural networks model

Contents	Pretreatment method	Number of hidden neurons	Transfer function	threshold value	Step size	Momentum	MSE	R^2_{cal}	RMSEP	R^2_{pre}
Aloe-emodin	SNV	8	TanhAxon	0.0001	0.8	0.6	0.0166	0.9580	0.0272	0.8902
	MMN	9	TanhAxon	0.01	0.9	0.7	0.0214	0.9289	0.0377	0.8326
	OE	11	TanhAxon	0.001	1.0	0.9	0.0130	0.9744	0.0371	0.8229
	Untreated	8	TanhAxon	0.001	0.5	0.9	0.0143	0.9715	0.1176	0.1591
Rhein	MMN	12	TanhAxon	0.005	0.7	0.9	0.0330	0.9850	0.0856	0.9080
	SNV	8	TanhAxon	0.0001	0.8	0.7	0.0292	0.9882	0.1039	0.8784
	OE	7	TanhAxon	0.005	0.9	0.8	0.0333	0.9850	0.1099	0.8593
	Untreated	7	TanhAxon	0.005	0.7	0.7	0.0334	0.9853	0.1126	0.8102
Emodin	SNV	12	TanhAxon	0.0001	0.8	0.9	0.0133	0.9937	0.0539	0.9187
	MMN	12	TanhAxon	0.0001	0.8	0.7	0.0196	0.9864	0.0679	0.8702
	Untreated	14	TanhAxon	0.001	1.0	0.8	0.0158	0.9912	0.0730	0.8475
	OE	11	TanhAxon	0.005	0.7	0.9	0.0249	0.9777	0.0842	0.7897
Chrysophanol	SNV	10	TanhAxon	0.0001	1.0	0.9	0.0556	0.9829	0.1813	0.8381
	MMN	7	TanhAxon	0.005	0.9	0.2	0.0899	0.9549	0.1640	0.8309
	Untreated	9	TanhAxon	0.001	0.6	0.9	0.0485	0.9896	0.1906	0.8069
	OE	8	TanhAxon	0.0001	0.9	0.9	0.0542	0.9837	0.2016	0.7970
Physcion	SNV	9	TanhAxon	0.001	0.9	0.9	0.0156	0.9878	0.0329	0.9419
	OE	5	TanhAxon	0.001	0.9	0.9	0.0187	0.9827	0.0532	0.8749
	Untreated	10	TanhAxon	0.0001	0.9	0.6	0.0154	0.9907	0.0528	0.8422
	MMN	8	TanhAxon	0.01	0.6	0.7	0.0395	0.9230	0.0532	0.8152

OE: Offset elimination; COE: Constant OE; MMN: Minimum/maximum normalization; SNV: Standard normal variate transformation; MSE: Mean square error; RMSEP: Root mean square error of prediction

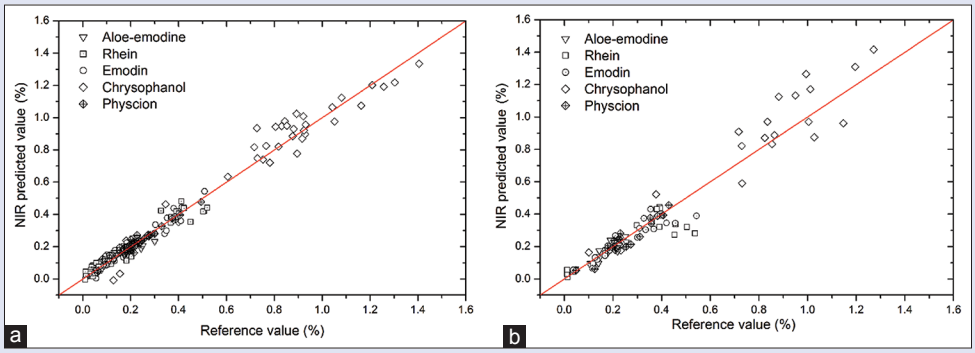


Figure 4: Comparisons between reference data and near infrared reflectance predicted values of the prediction set: (a) Partial least square model; (b) Artificial neural network model

Table 7: Results of the prediction set by the partial least squares and artificial neural networks model

Contents	Calibration method	RMSEP	R^2_{pre}
Aloe-emodin	PLS	0.0251	0.9161
	ANN	0.0272	0.8902
Rhein	PLS	0.0445	0.9699
	ANN	0.0856	0.9080
Emodin	PLS	0.0333	0.9655
	ANN	0.0539	0.9187
Chrysophanol	PLS	0.0862	0.9611
	ANN	0.1813	0.8381
Physcion	PLS	0.0211	0.9724
	ANN	0.0329	0.9419

PLS: Partial least squares; ANN: Artificial neural networks; RMSEP: Root mean square error of prediction

Result of near infrared reflectance models

The performance of PLS and Ann model was evaluated by the samples in prediction set. By comparing the values of RMSEP and R^2_{pre}, the relatively best-fitted calibration model of five free anthraquinones was finally obtained [Table 7]. The results of the *T*-test indicated no significant difference in the values of RMSEP and R^2_{pre} between the two types of calibration model. However, for aloe-emodin, rhein, emodin, chrysophanol, and physcion, both RMSEP and R^2_{pre} in the PLS models were superior to those in the ANN models, so PLS succeeded in developing optimum models with better fitting and prediction for the five free anthraquinones. Most of the R^2_{pre} in the PLS model were over 0.95, while the values of R^2_{pre} in the ANN model were less than 0.95. Figure 4 shows the results of the prediction set for the optimal PLS and ANN models. It is clearly that the PLS models have better prediction accuracy than the ANN models, especially for chrysophanol.

CONCLUSION

This study demonstrated that NIR spectroscopy together with PLS or ANN could be applied to determine the five free anthraquinones contents in rhubarb. The PLS and ANN calibration models of five free anthraquinones both had low RMSEP and high R^2_{pre} indicated that the 2 types of models were accurate, robust, and repeatable for rapid determination. For which this real-time measurement will significantly improve the efficiency of quality control and assurance of herbal medicine production. Compared to HPLC and LC-MS, this technique does not require tedious sample preparation, time-consuming sample analysis, or large amounts of solvent. It indicates that NIR is an ideal PAT tool that may be applied in the TCM industrial manufacturing process. Furthermore, as an analysis method that does not destroy samples, NIR is an indispensable means for sample quantification.

Financial support and sponsorship

This work was supported by the National Sciences Foundation of PR China (grant numbers 81373969); the University Key Research Projects of Henan Province (grant numbers 17A360026); the scientific and technological projects of Henan Province (grant numbers 172102310326); and the Cultivation Fund of Xinxiang Medical University (grant numbers 505095).

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. ICH Q8 (R2). Pharmaceutical Development: Part I: Pharmaceutical Development, Part II: Annex to Pharmaceutical Development. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q8_R1/Step4/Q8_R2_Guideline.pdf. [Last updated on 2009 Aug 20; Last accessed on 2018 Jul 13].
2. Streefland M, Martens DE, Beuvery EC, Wijffels RH. Process analytical technology (PAT) tools for the cultivation step in biopharmaceutical production. Eng Life Sci 2013;13:212-23.
3. Märk J, Andre M, Karner M, Huck CW. Prospects for multivariate classification of a pharmaceutical intermediate with near-infrared spectroscopy as a process analytical technology (PAT) production control supplement. Eur J Pharm Biopharm 2010;76:320-7.
4. National Pharmacopoeia Committee, China Food and Drug Administration. Pharmacopoeia of the People's Republic of China. Part 4. Beijing, China: Chinese Medical Science and Technology; 2015.
5. Jintao X, Yufei L, Liming Y, Chunyan L, Quanwei Y, Weiying W, *et al.* Rapid and simultaneous analysis of five alkaloids in four parts of *Coptidis rhizoma* by near-infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2018;188:611-8.
6. Feng Y, Lei D, Hu C. Rapid identification of illegal synthetic adulterants in herbal anti-diabetic medicines using near infrared spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2014;125:363-74.
7. Li W, Cheng Z, Wang Y, Qu H. Quality control of *Lonicerae japonicae* flos using near infrared spectroscopy and chemometrics. J Pharm Biomed Anal 2013;72:33-9.
8. Li K, Wang W, Liu Y, Jiang S, Huang G, Ye L. Near-infrared spectroscopy as a process analytical technology tool for monitoring the parching process of traditional Chinese medicine based on two kinds of chemical indicators. Pharmacogn Mag 2017;13:332-7.
9. Wang W, Xue J, Li K, Hu D. Dynamic predictive models of five alkaloids in Coptis during the process of stir-frying with wine using near-infrared spectroscopy. Int J Food Prop 2017;20 Supp 1:644-53.
10. Li W, Qu H. Characterization of herbal powder blends homogeneity using near-infrared spectroscopy. J Innov Opt Health Sci 2014;7:1-7.
11. Luypaert J, Massart DL, Vander Heyden Y. Near-infrared spectroscopy applications in pharmaceutical analysis. Talanta 2007;72:865-83.

12. Goodarzi M, Sharma S, Ramon H, Saeys W. Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. Trends Anal Chem 2015;67:147-58.
13. Kirchler CG, Pezzei CK, Beć KB, Mayr S, Ishigaki M, Ozaki Y, *et al.* Critical evaluation of spectral information of benchtop vs. portable near-infrared spectrometers: Quantum chemistry and two-dimensional correlation spectroscopy for a better understanding of PLS regression models of the rosmarinic acid content in *Rosmarini folium*. Analyst 2017;142:455-64.
14. Kirchler CG, Pezzei CK, Beć KB, Henn R, Ishigaki M, Ozaki Y, *et al.* Critical evaluation of NIR and ATR-IR spectroscopic quantifications of rosmarinic acid in *Rosmarini folium* supported by quantum chemical calculations. Planta Med 2017;83:1076-84.
15. Goodarzi M, Saeys W. Selection of the most informative near infrared spectroscopy wavebands for continuous glucose monitoring in human serum. Talanta 2016;146:155-65.
16. Collins RA. A ten-year audit of traditional Chinese medicine and other natural product research published in the Chinese Medical Journal (2000-2009). Chin Med J (Engl) 2011;124:1401-8.
17. Wang JB, Kong WJ, Wang HJ, Zhao HP, Xiao HY, Dai CM, *et al.* Toxic effects caused by rhubarb (*Rheum palmatum* L.) are reversed on immature and aged rats. J Ethnopharmacol 2011;134:216-20.
18. National Pharmacopoeia Committee, China Food and Drug Administration. Pharmacopoeia of the People's Republic of China. Part 1. Beijing, China: Chinese Medical Science and Technology; 2015. p. 23-4.
19. Ye D, Yuan S. A Dictionary of Processing Chinese Materia Medica. 1st ed. Shanghai: Scientific and Technical Publishers; 2005.
20. Chen LL, Verpoorte R, Yen HR, Peng WH, Cheng YC, Chao J, *et al.* Effects of processing adjuvants on traditional Chinese herbs. J Food Drug Anal 2018;26:S96-114.
21. Sun W, Zhang X, Zhang Z, Zhu R. Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb. Spectrochim Acta A Mol Biomol Spectrosc 2017;171:72-9.
22. Wang F, Zhang Z, Cui X, de B Harrington P. Identification of rhubarbs by using NIR spectrometry and temperature-constrained cascade correlation networks. Talanta 2006;70:1170-6.
23. Tang YF, Zhang ZY, Fan GQ. Identification of official rhubarb samples based on NIR spectra and neural networks. Guang Pu Xue Yu Guang Pu Fen Xi 2004;24:1348-51.
24. Ercan Ş, Uğur A. Artificial neural network models for lot-sizing problem: A case study. Neural Comput Appl 2013;22:1039-47.
25. Blanco M, Coello J, Iturriaga H, Maspoch S, Pages J. NIR calibration in non-linear systems: Different PLS approaches and artificial neural networks. Chemom Intell Lab 2000;50:75-82.
26. Falode O, Oghenerume S. Predicting the onset of asphaltene precipitation in heavy crude oil using artificial neural network. Chem Eng Process Res 2013;15:1-10.
27. Heigl N, Greiderer A, Petter CH, Kolomiets O, Siesler HW, Ulbricht M, *et al.* Simultaneous determination of the micro-, meso- and macropore size fractions of porous polymers by a combined use of Fourier transform near-infrared diffuse re flection spectroscopy and multivariate techniques. Anal Chem 2008;80:8493-500.
28. Wang P, Zhang H, Yang H, Nie L, Zang H. Rapid determination of major bioactive isoflavonoid compounds during the extraction process of kudzu (*Pueraria lobata*) by near-infrared transmission spectroscopy. Spectrochim Acta A Mol Biomol Spectrosc 2015;137:1403-8.
29. Chen QS, Zhao JW, Chaitep S, Guo ZM. Simultaneous analysis of main catechins contents in green tea (*Camellia sinensis* (L.)) by Fourier transform near infrared reflectance (FT-NIR) spectroscopy. Food Chem 2009;113:1272-7.
30. Xue J, Chen H, Xiong D, Huang G, Ai H, Liang Y, *et al.* Noninvasive measurement of glucose in artificial plasma with near-infrared and Raman spectroscopy. Appl Spectrosc 2014;68:428-33.
31. Chen Q, Qi S, Li H, Han X, Ouyang Q, Zhao J. Determination of rice syrup adulterant concentration in honey using three-dimensional fluorescence spectra and multivariate calibrations. Spectrochim Acta A Mol Biomol Spectrosc 2014;131:177-82.
32. Kachrimanis K, Rontogianni M, Malamataris S. Simultaneous quantitative analysis of mebendazole polymorphs A-C in powder mixtures by DRIFTS spectroscopy and ANN modeling. J Pharm Biomed Anal 2010;51:512-20.
33. Udelhoven T, Schutt B. Capability of feed-forward neural networks for a chemical evaluation of sediments with diffuse reflectance spectroscopy. Chemometr Intell Lab Syst 2000;51:9-22.