

ESTIMATION OF NICOTINE CONTENT IN TOBACCO LEAVES BASED ON HYPERSPECTRAL IMAGING

DOU, Y. Q.^{1,2} – CHENG, S.³ – LI, X. J.^{2*} – LIU, Y.⁴ – YUAN, X. L.¹

¹*Tobacco Research Institute, Chinese Academy of Agricultural Sciences
Qingdao, 266101, China*

²*College of Resources and Environment, Shandong Agricultural University
Taian, 271018, China*

³*Shanghai Tobacco Group Co., LTD, China Tobacco
Shanghai, 200082, China*

⁴*Shandong Linyi Tobacco Co., LTD, China Tobacco
Linyi, 276400, China*

**Corresponding author*

e-mail: lxj0911@126.com; phone: +13-5-05-388-080

(Received 10th Apr 2017; accepted 11th Aug 2017)

Abstract. We used spectroradiometry to rapidly and nondestructively predict the nicotine content in flue-cured tobacco leaves and obtained hyperspectral images of tobacco leaves during the squaring development stage at field conditions. We tested the nicotine content in the lab and analyzed correlations between 28 hyperspectral characteristic parameters and the nicotine content of tobacco leaves to establish inversion models. The result showed that $(SDr-Sy)/(SDr+Sy)$, Rg/Ro , and $(Rg-Ro)/(Rg+Ro)$ provided good correlations for the nicotine content; an inversion model based on $(SDr-Sy)/(SDr+Sy)$ in relation to the nicotine content, $Y = -4.2628 + 5.8974X_{(SDr-Sy)/(SDr+Sy)} - 1.3260X_{(SDr-Sy)/(SDr+Sy)}^2$, and a multiple regression model based on Rg/Ro , $(SDr-Sy)/(SDr+Sy)$ in relation to the nicotine content, $Y = -0.4753 + 1.6982X_{(SDr-Sy)/(SDr+Sy)} - 0.1616X_{Rg/Ro}$, could be used to predict the nicotine content of flue-cured tobacco leaves during the squaring stage.

Keywords: *flue-cured tobacco, spectral reflectance, chemical composition, inversion model, prediction*

Introduction

Tobacco was used in smoking products and nicotine was the most important chemical that influences the quality of tobacco, and it causes aphysiological dependence (Xu et al., 2003; Li et al., 2007). The nicotine content in tobacco leaves was an important indicator for the quality of the raw tobacco material; thus, it must be monitored to maintain a consistent and stable cigarette quality (Yan et al., 2001; Weeks et al., 1995).

It was a timely, convenient and nondestructive monitoring method to use hyperspectral reflectance to predict nicotine content in tobacco leaves (Xing et al., 2016). Near-infrared analysis was currently the most convenient way to detect the nicotine content after baking (Qiao, 2015). However, there was no monitoring method that can be applied to large tobacco fields. Many studies have examined the biochemical components and nutrient elements based on reflection spectral characteristics (Tang et al., 2003; Tong et al., 1997; Shibayama and Akiyama, 1989, 1991) and accordingly hyperspectral inversion prediction model was established. The results show that the optimum predicted method for nicotine content in tobacco leaves was the linear fitting of the two order differential of 1135nm reflectance (Li et al., 2006a). It was feasible to

predict the main chemical composition of a crop using hyperspectral images (Su et al., 2013; Zhao et al., 2014). However, most hyperspectral data was obtained in a controlled environment (Fang et al., 2013; Wang et al., 2011; Zhang and Zhao, 2000) hence, the inversion model and application effect have some limitations. So, there was still no effective and timely monitoring means for large areas of tobacco leaves. We selected a relatively concentrated producing area to examine the relationship between the hyperspectral parameter and the nicotine content in flue-cured tobacco under natural conditions. We set up an inversion model to predict the nicotine content in flue-cured tobacco leaves. We believe that the experiments will provide a good foundation for predicting nicotine content of large-forecast tobacco.

Materials and Methods

Study area

The study was carried out in a relatively concentrated flue-cured tobacco production area of Lin-yi, China. The study area was located in the northern part of Yishui county, Shandong Province (Latitude $35^{\circ}36' \sim 36^{\circ}13'$, longitude $118^{\circ}13' \sim 119^{\circ}03'$). The topography of study area was relatively unitary, and tobacco planting was carried out by the unified organization of the tobacco companies, unified management, unified planting, unified fertilization. Flue-cured tobacco varieties was NC102.

Sampling design

A total of 100 spatially isolated tobacco plants were sampled to determine their leaf-level nicotine content, of which 70 samples were used for modeling, and 30 samples were used. We identified individual tobacco sample locations by Gridding method. Depended on the green depth of tobacco leaves, locating these points in the field, and then selecting the nearest “suitable” tobacco plant. For the purpose of this study, the samples divided into three groups based on the color depth, which represents nitrogen under nutrition, normal nutrition, or nitrogen over nutrition. All data were recorded at midday (10:00–14:30, GMT+8) under clear skies and windless conditions.

Live tobacco measurements

Green-leaf samples were collected from each tobacco plant. Reflectance spectra (350–2500 nm) were collected for optically dense ground leaf samples using a high-intensity contact probe attachment (Analytical Spectral Devices (ASD) Inc., Boulder, CO, USA). The spectral resolutions of the instrument are 3 and 10 nm and the sampling intervals are 1.4 and 3.0 nm in the 350–1000 and 1000–2500 nm regions, respectively. The samples were killed at 105 °C for 15 min and then oven dried (60 °C for 72 h), passed through a 1-mm mesh screen, and weighed to the nearest 0.1 mg. The leaf-level nicotine concentrations (grams of N/100 g sample) of the oven dried ground foliage were determined using a LECO TruSpec CN analyzer (St. Joseph, MI, USA). For each sample, a series of twenty replicates were obtained, and the probe was then re-calibrated using a white reflectance cap accessory available from the manufacturer.

Spectral data processing

Leaf-level reflectance data were transformed using standard derivative analysis and continuum removal methodologies (Xiao, 1998). Both transformations were applied directly to the spectrometer data and/or to a smoothed version of the spectrometer data. We opted to smooth the spectrometer data by resampling using coarser HyMap sensor channel configurations (472~2488 nm; 15~20 μm bandpass and sampling interval) because we are familiar with the sensor and the HyMap imagery has been successfully used for remote sensing of foliar biochemistry. Standard derivative analysis was implemented by calculating first derivative reflectance (FDR) values for log-transformed reflectance data ($\log_{10}(1/R)$, where R was reflectance) using the ViewSpec Pro software (version 6.0, ASD). The hyperspectral indices were chosen according to previous studies (Huang et al., 2014; Liang et al., 2013; Li et al., 2006b; Deng et al., 2008; Wang et al., 2011; Xia et al., 2009; Hu et al., 2001; Breuer and Pavan, 1955; Nevalainen et al., 2014; Li et al., 2013; Yang et al., 2011) and are presented in *Table 1*. The related coefficients between the nicotine content and vegetation index were calculated using Data Processing System (DPS) software (Canfield Clinical Systems, Fairfield, New Jersey). A model fitting effect was tested using the relative error (RE), determination coefficient (R^2), and root-mean-square error (RMSE).

Results

Correlation analysis for tobacco leaf nicotine content and high spectral index

A significant correlation was observed between the six indexes of the tobacco leaves and the nicotine content (*Table 1*). The correlation coefficients for five hyperspectral indices and the nicotine content of tobacco leaves were extremely significant. The absolute value of the correlation coefficient between the $(\text{SDr}-\text{Sy})/(\text{SDr}+\text{Sy})$ index and the tobacco leaf nicotine content was above 0.80. The absolute values of Rg/Ro , $(\text{Rg}-\text{Ro})/(\text{Rg}+\text{Ro})$, and the nicotine content of the tobacco leaves were all above 0.40.

Table 1. The correlation coefficients between the nicotine content of tobacco leaves and the spectral parameters ($n=70$)

Index	Correlation Coefficient	Index	Correlation Coefficient
Db	-0.1677	SDr/Sy	-0.1930
λ_b	0.2273 ^a	SDr/SDg	0.1621
SDb	-0.2481 ^a	$(\text{Rg}-\text{Ro})/(\text{Rg}+\text{Ro})$	-0.4014 ^b
Dy	0.2550 ^a	$(\text{SDr}-\text{SDb})/(\text{SDr}+\text{SDb})$	0.3080 ^a
λ_y	-0.010	$(\text{SDr}-\text{Sy})/(\text{SDr}+\text{Sy})$	0.8707 ^b
Sy	0.1838	$(\text{SDr}-\text{SDg})/(\text{SDr}+\text{SDg})$	0.1650
Rg	-0.1371	NDSI(1350, 700)	0.2043 ^a
SDg	-0.1316	NDSI(FD700, FD690)	0.3496 ^b
Ro	0.0978	NDVI(573, 440)	-0.3980 ^b
Rg/Ro	-0.4152 ^b	NDVI(660, 440)	-0.2986 ^a
SDr/SDb	0.2804	NDVI(1220, 610)	0.2081 ^a

^a $P \leq 0.05$, ^b $P \leq 0.01$

Establishment and testing of the nicotine content estimation model for tobacco leaves

The parameters for the correlation coefficient of the nicotine content in different growth stages for the tobacco leaves (independent variables) were extremely significant. A linear function ($y=a+bx$), power function ($y=ax^b$), logarithmic function ($y=a+b \ln(x)$), quadratic function ($y=a+bx+cx^2$), exponential function ($y=ae^{bx}$), and negative exponential function ($y=a/(1+\exp(b+cx))$) were chosen as mold functions to establish a tobacco leaf nicotine content inversion model (where y represents the nicotine content of the tobacco leaves and x represents the hyperspectral eigen values; a , b , and c are constants). Then, the Root Mean Square Error (RMSE) and Relative Error (RE) were selected to evaluate the fitting effect between the estimated value and the measured value to verify the accuracy and applicability of the model. Two hyperspectral indices, Rg/Ro and $(SDr-Sy)/(SDr+Sy)$, with high correlation coefficients were chosen as the independent variables for the estimation model to establish the univariate estimation model and multiple regression estimation model.

Establishment and testing of the univariate estimation model for nicotine content in tobacco leaves

The inversion model was established using the hyperspectral indices, Rg/Ro and $(SDr-Sy)/(SDr+Sy)$, as independent variables and nicotine content as the dependent variable (Table 2).

Table 2. Nicotine concentration retrieval model for tobacco leaves during the squaring development stage and the test results

Index	Estimation model	R^2 (n=70)
$(SDr-Sy)/(SDr+Sy)$	$y = -1.1977 + 1.7989x$	0.7582**
	$y = 2.2427 / (1 + \exp(6.1463 - 4.7476x))$	0.7955**
	$y = 0.7597x^{1.6421}$	0.7210**
	$y = 0.3258 \times \exp(0.9926x)$	0.6667**
	$y = -4.2628 + 5.8974x - 1.3260x^2$	0.8112**
Rg/Ro	$y = 3.0047 - 0.4399x$	0.1724
	$y = 2.1130 / (1 + \exp(-4.5275 + 1.0456x))$	0.1784
	$y = 4.7454x^{-0.9553}$	0.1557
	$y = 4.0603 \times \exp(-0.2945x)$	0.1669
	$y = 2.3793 - 0.0833x - 0.0498x^2$	0.1734

From the data in Table 2, the decisive coefficients of the univariate inversion model (Eq. 1) and (Eq. 2) with the hyperspectral parameter $(SDr-Sy)/(SDr+Sy)$ as the independent variable were all significant.

$$Y = 2.2427 / (1 + \exp(6.1463 - 4.7476X_{(SDr-Sy)/(SDr+Sy)})) \quad (\text{Eq. 1})$$

$$Y = -4.2628 + 5.8974X_{(SDr-Sy)/(SDr+Sy)} - 1.3260X_{(SDr-Sy)/(SDr+Sy)}^2 \quad (\text{Eq. 2})$$

The estimation model was tested using the nicotine content data from 30 tobacco leaf samples, and the estimated RMSE and RE (%) were used to evaluate the fitting effect of

the model. The R^2 value was high and the RMSE and RE (%) values were small for the inversion model (Eq. 1) and (Eq. 2).

Establishment and test of the multiple regression model for nicotine content in tobacco leaves

The multiple regression model was established using R_g/R_o and $(SDr-Sy)/(SDr+Sy)$ models (with higher correlation coefficients) as the independent variables. The regression equation was (Eq. 3).

$$Y = (-0.4753 + 1.6982X_{(SDr-Sy)/(SDr+Sy)} - 0.1616X_{R_g/R_o}) \quad (\text{Eq. 3})$$

$F = 135.7792$, $p = 0.0000$, $R = 0.8827$, and $R^2 = 0.7791$ (adjusted correlation coefficient, $R_a = 0.8794$; adjusted coefficient of determination, $R_a^2 = 0.7734$). The nicotine content of the tobacco leaves was predicted using the model (Eq. 3) and the correlation coefficient between the predicted value and the measured value was 0.8841; RMSE = 0.2335 for the prediction equation and RE = 15.47%.

Optimum estimation model of nicotine content in tobacco leaves

Single-variable prediction model

Based on the above analysis, the nicotine content of the tobacco leaves from the single-variable prediction model was Eq.2.

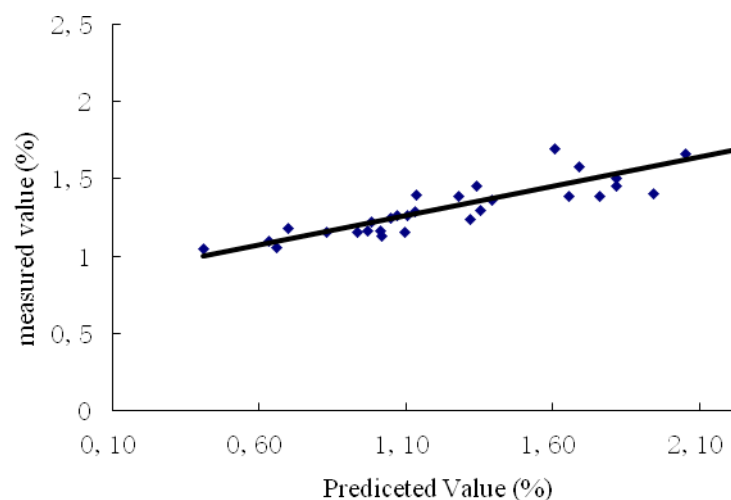


Figure 1. Validation result of estimation model of nicotine content of tobacco

The estimated RMSE and RE (%) were used to evaluate the fitting effect of the model. The correlation coefficient of the fitted equation was 0.8148. The RMSE of the fitted equation was 0.2272 and the RE was 14.42%. Further tests for the accuracy of the model (Eq. 2), provided the measured nicotine value and the estimated value of the 1:1 diagram. The correlation coefficient between the predicted value and the measured value was 0.7894. The model provides good precision and can be used to predict the nicotine content in tobacco leaves.

Multivariate prediction model

Based on the above analysis, the nicotine content of tobacco leaves from the multivariate prediction model was Eq.3.

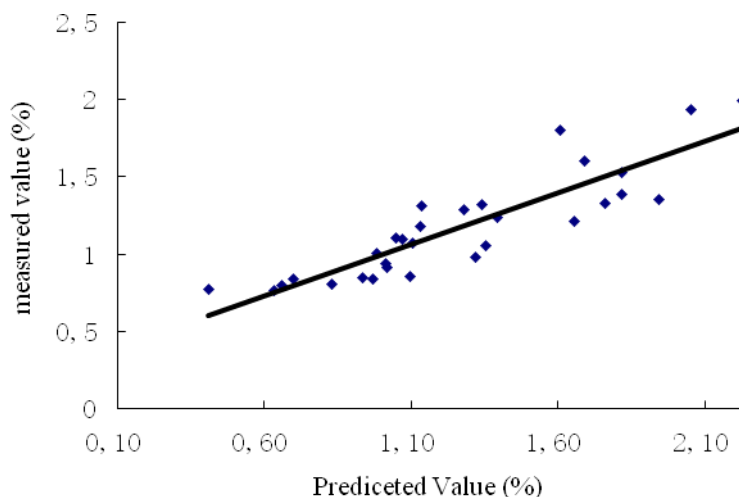


Figure 2. Validation result of estimation model of nicotine content of tobacco

The estimated RMSE and RE (%) were used to evaluate the fitting effect of the model. The correlation coefficient of the fitted equation was 0.8841. The RMSE of the fitted equation was 0.2335 and the RE was 15.47%. Further tests for the accuracy of the model (Eq. 3), provided the measured nicotine value and the estimated value of the 1:1 diagram. The correlation coefficient between the predicted value and the measured value was 0.7816. The model provides good precision and can be used to predict the nicotine content in tobacco leaves.

Discussion

Although there are technical standards and uniform requirements for tobacco cultivation in China, annual and regional differences of the nicotine content are significant because of the effects of natural precipitation, soil fertility, and field management measures. Nicotine was the most important tobacco leaf quality factor, and its content determines the intrinsic quality of tobacco leaves and the status of the formula. Industrial enterprises need to predict the main chemical components (especially nicotine content) of tobacco leaves before tobacco procurement, and reduce annual chemical composition differences of tobacco leaves by allocating different tobacco production plans to maintain a stable quality for tobacco brands. This study could only predict the nicotine content of fresh tobacco leaves by screening hyperspectral parameters, establishing models, and optimizing the validation methods. However, we did not establish the relationship between the spectral characteristic parameters and the nicotine content of the initial flue-cured tobacco after preparation. The method could not provide a practical basis for the changes of the cigarette tobacco leaf and recipe directly from the operational level. Future work would include exploring

the precise labeling of hyperspectral tobacco leaves, studying the nicotine content of the first flue-cured tobacco, establishing an estimation model, and predicting the nicotine content of the early flue-cured tobacco to provide the basis for tobacco leaf dispensing and formulation allocation to the tobacco industry.

Conclusion

Using tobacco plants with different nicotine contents growing in natural field conditions, we analysed its hyperspectral characteristics, the Correlation between hyperspectral parameters and nicotine content. Hyperspectral parameters related to nicotine content in tobacco leaves have been preliminarily established and A prediction model of nicotine content of tobacco leaves with high precision was established. The results show that there was a high correlation between the tobacco nicotine content and the $(SDr - Sy)/(SDr + Sy)$, Rg/Ro , and $(Rg - Ro)/(Rg + Ro)$ parameters. The equations for the univariate estimation models (1) and (2), can be used to predict the nicotine content of flue-cured tobacco leaves. The multivariate regressive estimation model equation (3) was established using hyperspectral parameters Rg/Ro and $(SDr - Sy)/(SDr + Sy)$ and can be used to predict the nicotine content in flue-cured tobacco leaves. The results of this study and the estimation model are closer to the actual situation for the flue-cured tobacco growth and development environment, and can provide a reference for the prediction of tobacco nicotine content in large areas.

Acknowledgements. Supported by National Natural Science Foundation of China (No:41171425) and by China National Tobacco Corp project (No: 110201402007).

REFERENCES

- [1] Breuer, M. E., Pavan, C. (1955): Behaviour of polytene chromosomes of *Rhynchosciara angela* at different stages of larval development. – *Chromosoma* 7: 371-386.
- [2] Deng, X. X., Zhou, J. H., Chen, X. L. et al. (2008): Correlation analysis on quality evaluating indexes in tobacco leaf. – *Acta Tabacaria Sinica* 14(2):1-8.
- [3] Fang, X. Y., Zhu, X. C., Wang, L. et al. (2013): Hyperspectral monitoring of the canopy chlorophyll content at apple tree prosperous fruit stage. – *Scientia Agricultura Sinica* 4(16): 3504-3513.
- [4] Hu, J. J., Ma, M., Li, Y. G. et al. (2001): Grey Incidence analysis on the correlation between main chemical components and sensory quality of flue-cured tobacco. – *Tobacco Science & Technology* 152(1):3-7.
- [5] Huang, W., Yang, Q., Pu, R. et al. (2014): Estimation of nitrogen vertical distribution by bi-directional canopy reflectance in winter wheat. – *Sensors*, 14(11):20347-20359.
- [6] Li, C., Li, H. Z., Liu, G. S. (2007): Advances in research on the optimization of nicotine conversion and alkaloid in tobacco. – *Journal of Henan Agricultural Sciences* (6):14-17.
- [7] Li, F. L., Zhao, C. J., Liu, L. Y., Cao, W. X. (2006a): Screening of hyper-spectral estimation method on biochemical components of fresh flue-cured tobacco leaves. – *Transactions of The Chinese Society of Agricultural Engineering* 22(3): 88-94.
- [8] Li, F. Z., Feng, M. C., Yang, W. D. et al. (2013): Monitoring of winter wheat chlorophyll content in irrigated and dry lands of Shanxi Prov-ince of China based on hyperspectral remote sensing. – *Chinese Journal Ecology* 32(12):3213-3218.

- [9] Li, J. J., Pang, T. H., Liu, G. S. et al. (2006b): Correlation and path analysis on smoking aroma quality and chemical components in flue-cured tobacco. – *Agricultural Science Bulletin* 22(1): 71-73.
- [10] Liang, J., Meng, K. L., Tian, M. M. et al. (2013): Diagnosis of nitrogen content in upper and lower corn leaves based on hyperspectral data. – *Spectroscopy and Spectral Analysis* 33(4):1032-1037.
- [11] Nevalainen, O., Hakala, T., Suomalainen, J. et al. (2014): Fast and nondestructive method for leaf level chlorophyll estimation using hyperspectral LiDAR. – *Agricultural & Forest Meteorology* 198-199:250-258.
- [12] Qiao, X. (2015): Based on Nir-master for the conventional chemical composition analysis model of flue-cured tobacco. – *YANBIAN University*, 5-6.
- [13] Shibayama, M., Akiyama, T. (1989): Seasonal visible, near-infrared and mid-ground dry phytomass. – *Remote Sensing of Environment* 27:119-127.
- [14] Shibayama, M., Akiyama, T. (1991): Estimating grain yield of maturing rice canopies using high spectral resolution reflectance measurements. – *Remote Sensing of Environment* 36:45-53.
- [15] Su, Y. Sh., Su, F. Q., Wang, H. B. et al. (2013): Relationship between tobacco canopy spectral features and chlorophyll, leaf area index under different fertilization conditions. – *Chinese Tobacco Science* 34(2): 23-27.
- [16] Tang, Y. L., Wang, J. H., Hang, J. F. (2003): Study of change law between hyperspectrum and chlorophyll, carotenoid in ripening process of rice. – *Transactions of The Chinese Society of Agricultural Engineering* 19: 167-173.
- [17] Tong, Q. X., Zheng, L. F., Wang, J. N. et al. (1997): Study on imaging spectrometer remote sensing information for wetland vegetation. – *Journal of Remote Sensing* 1(1): 50-57.
- [18] Wang, K. R., Pan, W. Ch., Li Sh. K. et al. (2011): Monitoring Models of the Plant Nitrogen Content Based on Cotton Canopy. – *Hyperspectral Reflectance.Spectroscopy and Spectral Analysis* 31(7):1868-1872.
- [19] Weeks, W. W., Campos, M. P., Moldoveanu, S. (1995): Pyrolysis of cherry red tobacco and 1-deoxy-1-[(S)-2-(3-pyridyl)-l-pyrrolidinyl]-.bet.-D-fructose(pyranoose and furanoose isomers) amadori products of cherry red tobacco. – *Agric. Food Chem.* 43(8): 2247-2253.
- [20] Xia, Y., Yan, Z. H., Wei, F. et al. (2009): Exploring Novel Hyperspectral Band and Key Index for Leaf Nitrogen. Accumulation in Wheat. – *Hyperspectral Reflectance.Spectroscopy and Spectral Analysis* 29(8):2191-2195.
- [21] Xiao, X. Z. (1998): Tobacco Chemical. – Beijing, China Agriculture Press, 162, 205.
- [22] Xing, X. X, Zhang, T., Liu, B. Y. (2016): Estimation model of nicotine content in flue-cured Tobacco based on hyperspectra. – *Shandong Agricultural Sciences* 48(6): 142-146.
- [23] Xu, Y. M., Wang, S. S., Lai, L.,X. et al. (2003): Review on the studies of tobacco alkaloids. – *Chinese Tobacco Science* (2): 12-16.
- [24] Yan, K. Y., Wang, J. M., Qu, J. B. et al. (2001): Correlation between smoking quality and physical/chemical parameters of flue -cured tobacco in Henan Provinc. – *Tobacco Science & Technology* (1): 5-9.
- [25] Yang, X. H., Huang, J. F. et al. (2011): Estimating biophysical parameters of rice with remote sensing data using support vector machines. – *Science China (Life Science)*54(3): 272-281.
- [26] Zhang, Q. Y., Zhao, C. J. (2000): Mechanism analysis of leaf biochemical concentration by high spectral remote sensing. – *Remote Sensing of Environment* 4(2): 125-129.
- [27] Zhao, W., Liu, G. S., Jia, F. F. et al. (2014): Hyperspectral prediction model of nicotine content in flue cured tobacco. – *Jiangsu Agricultural Sciences* 42(3):275-279.