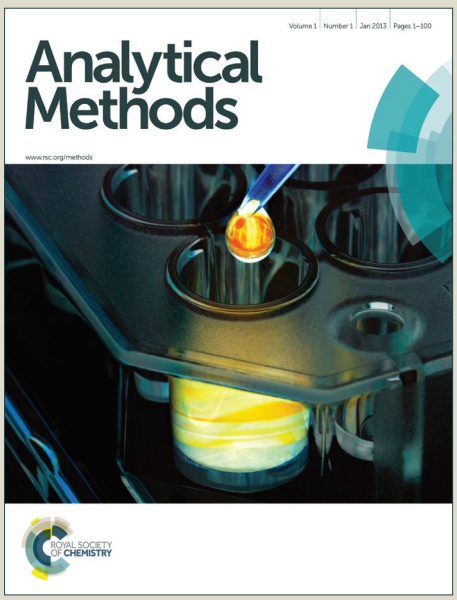
# Analytical Methods

Accepted Manuscript




This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

# A rapid and novel method for predicting nicotine alkaloids in tobacco through electronic nose and partial least square regression analysis

**Shunshun Lin[a], Xiaoming Zhang[a,*]**

[a] State Key Laboratory of Food Science and Technology, School of Food Science and Technology, Jiangnan University, Wuxi, 214122 Jiangsu, PR China;

## Author information

Corresponding author: Prof. Xiaoming Zhang.

Post address: State Key Laboratory of food Science and Technology, School of Food Science and Technology, Jiangnan University, Lihu Road 1800, Wuxi, Jiangsu 214122, China.

Tel: +86-510-85919106; Fax: +86-510-85884496. E-mail: xmzhang@jiangnan.edu.cn

## Abbreviations used

PLSR, partial least squares regression; B, the upper leaves of tobacco; C, the middle leaves of tobacco; X, the lower leaves of tobacco; correlation coefficient (R); regression coefficient ($R^2$); root mean square error of prediction (RMSEP); Gas chromatography-triple quadrupole mass spectrometry (GC-TriQ-MS).

**Abstract:** Alkaloid levels in tobacco are of great concern due to nicotine addiction and associated diseases. A rapid method for analyzing tobacco alkaloids is required for legislatures and tobacco companies. This study aims to establish prediction models of tobacco alkaloids through electronic nose responses and partial least squares regression (PLSR) for rapid analyzing alkaloids level in tobacco. Eight alkaloids (nicotine, myosmine, etc) were detected through gas chromatography-triple quadrupole mass spectrometry (GC-TriQ-MS). Characterization of alkaloids in different leaf positions (upper (B), middle (C) and lower (X)) was investigated and three signal features of electronic nose sensors were extracted for better modeling. Results showed that total alkaloid content significantly varied in the following order B>C>X. Sensors' maximum intensity ($IN_{max}$) and slope (K) were significantly related to alkaloids' level. Prediction models of alkaloids were successfully established. The calibrated ($R_{\_cal}$ of 0.99, $R^2_{\_cal}$ of 0.98) and validated ($R_{\_val}$ of 0.97, $R^2_{\_val}$ of 0.94) parameters for nicotine prediction model were very satisfactory. After validity checking, the established model for nicotine detection has 96% of prediction capability. Moreover, the prediction effectiveness of other alkaloids' models (except nicotyrine) was also proved accurate. This work provided evidence that electronic nose could be used as a testing tool to rapidly and quantitatively detect the content of nicotine alkaloids in tobacco. Further study is still needed to improve the precision and robustness of the alkaloids calibration models.

**Keywords:** Prediction; mathematical model; electronic nose; tobacco alkaloids; partial least squares regression (PLSR).

## 1. Introduction

Nicotine is regarded as the most serious health hazardous component in tobacco, accounting for over 95% of total tobacco alkaloids [1]. It does not only induce cigarette addiction, but also responsible for smoke associated diseases, due to the formation of tobacco specific nitrosamines (TSNAs) that were formed from the nitrosation of nicotine and related alkaloids during tobacco aging, curing and burning [1, 2]. Furthermore, although the minor alkaloids (myosmine, nornicotine, anabasine, nicotyrine, anatabine, 2,3-dipyridyl and cotinine, etc) existed in low level, they play an important role in smoking addiction, particularly myosmine and anatabine could increase the desire for nicotine, thus enhance smoking behavior [3].

With the enactment of the Family Smoking Prevention and Tobacco Control Act (FSPTCA), the U.S. Food and Drug Administration (FDA) encouraged to reduce nicotine to levels that are not addictive for protecting public health. It is consistent with the relevant articles of World Health Organization Framework Convention on Tobacco Control (FCTC), in which, that allow governmental agencies to establish standards for nicotine [4]. Moreover, governments and public health authorities in various parts of the world considered that lower nicotine yielding cigarette is an effective approach to reduce health risks of smoking from temporary "smoking reduction" to potentially permanent "smoking cessation" [5]. Therefore, a rapid and convenient method for controlling tobacco alkaloids level is required for legislatures and tobacco companies.

Moreover, variety, soil and leaf position on the plant are all among the variables

that influence tobacco grade and acceptability [6-8]. Tobacco leaves in China were mainly classified according to leaf positions (upper (B), middle (C) and lower (X) leaves), and different leaf positions indicate different quality grades. Sun, *et al*. clearly demonstrated that there have significant differences on neutral volatiles levels in different leaf positions of tobacco [8]. However, there have been limited reports on the characterization of alkaloids of flue-cured tobacco from different leaf positions.

A number of gas chromatography-mass spectrometry (GC/MS) [1, 9] and liquid chromatography-tandem mass spectrometry (LC-MS/MS) [10] methods were performed to identify tobacco alkaloids. However, these methods need tedious extraction before analysis, which inhibited analysis efficiency. Compared to GC/MS and LC-MS/MS, electronic nose systems are convenient, rapid and useful for both laboratory and industrial production field [11], they were widely applied in the food control principally for recognition and classification [12], such as electronic nose can distinguish or differentiate the freshness of beef strip loins samples [13], varieties of different rough rice samples [14], characteristic aroma of Chinese famous liquors [15], counterfeit for different tobacco brands [16], and quality of different oranges and apples [17]. Moreover, electronic nose systems have been successfully used to distinguish different cigarettes [16, 18].

Based on eighteen metal-oxide semiconductor sensors, electronic nose (Fox 4000 nose) changes its electrical resistances of sensors when these sensors were exposed to volatile substances, thus generating analytical signal [19, 20]. It was widely used to establish prediction model, such as, to assess the harvest season of peach [21], to predict

the chemical parameters of controlled oxidation tallow[22], and to evaluate the sensory quality of pork [20]. However, no studies exist so far on the use of gas sensor arrays to predict tobacco alkaloids.

Partial least square regression (PLSR) focuses on a comprehensive evaluation of information obtained from the raw data, and has been effectively used to explain the correlation of variables through reducing the dimensionality of the raw data set without losing information [23-25]. It is an effective tool to deal with multiple linear regression (MLR) problems: limited number of observations, missing data and collinearity [26].

The objective of this study was to establish an efficient tobacco alkaloids controlling method. Meanwhile, a convenient identification procedure for tobacco alkaloids was described. This study would pave a way to better control tobacco quality through supervising the level of tobacco alkaloids.

## 2. Experimental

### 2.1. Experimental materials and reagents

Forty-two flue-cured tobacco samples of "Yunyan 87" cultivar (2010) sourced from fourteen origins and three leaf positions (upper (B), middle (C) and lower (X) leaves) were used during this work. These samples were divided into two groups, the first group contained twenty-four samples, obtained from eight origins (followed as: Shaoyang, Longhui and Chenzhou City of Hunan Province, Xingyi and Zhengan City of Guizhou Province, Changning and Wenshan City of Yunnan Province, and Fengjie

city of Chongqing Municipality) and three leaf positions (B, C and X), were used for tobacco alkaloids' characterization analysis and model calibration/optimization. The second independent group contained eighteen samples, originated from six origins (followed as: Yongxing and Panxian City of Hunan Province, Zunyi City of Guizhou Province, Baoshan and Anning City of Yunnan Province, and Yuqing City of Chongqing Municipality) and three leaf positions, were used for model validation. These samples were dried at 35 $^{o}$C for 24 h, then they were ground and sieved through a 100 mesh screen for GC/MS analysis and electronic nose analysis.

Dichloromethane (HPLC grade) and methanol (HPLC grade) were purchased from Merck (Darmstadt, Germany). Alkaloids standards of nicotine, nornicotine, anabasine and anatabine, quinoline (99.9%) and n-alkanes (C8～C40) were purchased from Sigma-Aldrich (Shanghai, China). Sodium chloride, sodium hydroxide, and other chemicals were purchased from Sinopharm Chemical Reagent Co., Ltd. (Shanghai, China).

## 2.2. GC-MS analysis of alkaloids

Sample preparation process was adopted as published by Cai, et al. [27] with some modifications. About 0.400 g tobacco powder and 2.5 mL of 5% (g/g) NaOH solution were placed into a 50 mL plastic screw-capped tube. Then, 10.00 mL of 50 g mL$^{-1}$ quinoline extract liquor (dichloromethane: methanol=3:1) was added to the tube and mixture was ultrasonicated for 30 min at 20 $^{o}$C. Finally, about 2 mL extract solution (the lower solution) was taken and dehydrated with anhydrous sodium sulfate. The solution was filtered with a 0.22 μm filter membrane and stored in a 1.5 mL screw

capped vial for analysis. Extraction of each sample was performed in triplicate.

Gas chromatography-triple quadrupole mass spectrometry (GC-TriQ-MS) analysis was performed on TSQ Quantum XLS system from Thermo Fisher Scientific Inc (USA). Alkaloids analysis was carried out by DB-5MS column (30 m×0.25 mm×0.25 μm) in SIM scan mode according to the method of Lisko, *et al* [1] and Cai, et al. [27]. The temperature of column was programmed from 110 $^{o}$C (hold for 1 min) to 180 $^{o}$C (hold for 1 min) with a rate of 4 $^{o}$C min$^{-1}$, then raised to 250 $^{o}$C (held for 1 min) with a rate of 20 $^{o}$C min$^{-1}$. Helium was used as a carrier gas at a constant flow rate of 1 mL min$^{-1}$. The mass range was scanned from 45 to 350 m/z at 0.2 s/s can for the full-scan mode. The mass spectrometer was operated in the electron ionization (EI) mode at 70 eV.

A series of n-alkanes (C8～C40) was analyzed under the same conditions to get the linear retention index (RI). RI was calculated following the formula:

$$RI = 100n + \frac{100 \times [TR(x) - TR(n)]}{[TR(n+1) - TR(n)]}$$

Where TR is the retention time, $n$ and $n$+1 are the number of carbon in the alkanes eluting before and after the component $x$, respectively.

## 2. 3. Electronic nose analysis

Fox 4000 nose sensor (Alpha M.O.S. Toulouse, France) equipped with eighteen metal oxide sensors and a headspace auto-sampler HS100 were employed to analyze the volatile substances from tobaccos. In this study, the eighteen sensors were numbered as: (1): LY2/LG, (2): LY2/G, (3): LY2/AA, (4): LY2/GH, (5): LY2/gCTl, (6): LY2/gCT, (7): T30/1, (8): P10/1, (9): P10/2, (10): P40/1, (11): T70/2, (12): PA/2, (13):

P30/1, (14): P40/2, (15): P30/2, (16): T40/2, (17): T40/1 and (18): TA/2. The method was adopted from the published study by Song et al. [22] with some modifications: about 0.400 g sample powder was transferred to 10 mL glass vials with preheated Teflon/silicon septa and screw capped. Then the vials were placed in the auto-sampler of electronic nose. The temperature program of headspace was: after the samples were incubated at 60 $^{o}$C for 10 min, a headspace gas was pumped into the sensor chamber for 10 s at a flow rate of 150 mL min$^{-1}$. The recovery time was 120 s and the maximum resistance changes of each sensor were used for analysis to simplify the data processing [22, 28]. Each sample was analyzed for four times, and the average result was used for prediction analysis for getting stable result.

2. 4. Data analysis

The mean ± standard deviation (SD) content of tobacco alkaloids was calculated by analysis of variance (ANOVA) ($P < 0.05$) (SPSS 13.0 , Corporation, USA). Principal component analysis (PCA) and partial least squares regression (PLSR) analysis were carried out by Unscrambler version 9.7 (CAMO ASA, Oslo, Norway). The characterization of sensor responses of tobacco in different leaf positions and the correlation between tobacco alkaloids and sensor responses were analyzed by PCA and PLS2 method (PLSR was performed by many X-variables and several Y-variables simultaneously), respectively. Model calibration/optimization was performed through PLS1 (PLSR was performed by many X-variables and only one Y-variable) on jack-knifing test. All variables were centered and standardized (1/Sdev) for getting unbiased contribution of each variable to the criterion [22]. The significance was at $P <$

*0.05* level.

## 3. Results and discussion

### 3.1. GC/MS analysis of nicotine alkaloids in tobacco

Eight alkaloids in tobacco leaves were identified by mass spectrum and retention index (RI or Kovats index) in accordance with the authentic standard compositions (Fig.1) and literature reference data (shown in Table 1). Results were considered trusted since the differences between measured RI values (MRI) and referenced RI values were less than 10 (Table 1) [29]. Meanwhile, selected ion scanning module (SIM) was used for quantitative analysis because the content of the minor alkaloids is extremely lower than nicotine [1]. The quantitative ions of eight alkaloids were selected through mass spectrum analysis, and separately scanned in different time segments, that have been described in Table 1. Overall, eight structurally related alkaloids including nicotine, myosmine, nornicotine, anabasine, nicotyrine, anatabine, 2,3-dipyridyl and cotinine were precisely determined.

### 3.2. Characterization of tobacco alkaloids in different leaf positions

From Table 1, it was observed that there was a significant difference in the content of tobacco alkaloids from different leaf positions. The content of most alkaloids was significantly ($P<0.05$) higher in upper leaves (B) than middle (C) and lower (X) parts of the leaves, except for myosmine, nicotyrine, 2,3-dipyridyl and cotinine which showed non-significant difference between B and C parts of the leaves. Meanwhile, the content of nicotine, nornicotine, nicotyrine, 2,3-dipyridyl and cotinine

in C parts of the leaves was significantly ($P < 0.05$) higher than in X part of the leaves. Among these alkaloids, nicotine contributed greatly to total alkaloid content, since it is the most abundant alkaloid in tobacco [1] . Therefore, the level of total alkaloids in tobacco significantly followed the order B>C>X. The possible reason might be that the sunshine could easily reach the upper part of the leaves (B) than other parts, thus accelerate the transformation and absorption of nitrogen [30] and lead to higher nicotine alkaloids.

### 3.3. Sensor signal feature extraction

Fig. 2 shows the sensor signals of typical flue-cured tobacco. The intensity of each sensor is given in units of $(R_o-R_t)/R_o$, where $R_o$ was sensor's electrical resistance of detecting clean air (at t=0), $R_t$ was the electrical resistance in detecting process. The intensity has been expressed as conductivity in previous studies [21, 31]. From Fig. 2, it can be seen that the intensity of all sensors initially increased and subsequently decreased afterward.

As shown in Fig. 2, the feature of maximum intensity ($IN_{max}$) reflects the maximum concentration of volatile substances received by sensors during electronic analysis. The other features like slope and $T_{max}$ might be related to the volatility of analyzed substances. In present study, each sensor's three signal features contained the maximum intensity ($IN_{max}$), slope (K) and the time where the maximum intensity occurred ($T_{max}$). All these data were extracted for prediction assessment of tobacco alkaloids to avoid missing relevant additional information of analytes. The correlation between these signal features responses and tobacco alkaloids contents was

investigated through PLSR analysis. As a result, it was found that the responses of $T_{max}$ feature showed weak correlation with alkaloids contents. However, the responses of maximum intensity ($IN_{max}$) and slope (K) features were significantly correlated to the alkaloids content. And $IN_{max}$ responses of sixteen sensors and K responses of ten sensors, as listed in Table 2, were the main contributors to the establishment of alkaloids prediction models.

### 3.4. Characterization of sensor responses for Flue-cured tobacco

The significant signal features of $IN_{max}$ and K (listed in Table 2) were used for further analysis of the difference of tobacco leaves from different positions (B, C and X). The score plot of these feature responses (Fig. 3) by principal component analysis (PCA) explained 70% of the variance in PC1 and 10% of variance in PC2. The distance between the points on the plot reflects the difference among samples. From Fig. 3, it can be seen that samples were obviously divided into three groups. It was observed that upper-leaf samples (B1~B8) were located in the right part of the plot, middle-leaf samples (C1~C8) were situated closer to the center, and lower-leaf samples (X1~X8) were located in the left part of the plot. The overall difference of tobaccos in different leaf positions was distributed in the sequence of X, C, B along $PC_1$ from left to right, which is in good agreement with the order of nicotine alkaloids content in tobaccos. These results indicate that electronic nose system could be useful for analysis of nicotine alkaloids level in tobacco leaves. Similar researches have demonstrated that electronic nose systems could be successfully used to distinguish different cigarettes [28, 32].

### 3.5. Correlation between tobacco alkaloids and sensor responses

The derived PLSR model for sensor responses of above signal features (X-matrix) and tobacco alkaloids content (Y-matrix) included three significant PCs explaining 83% of cross-validated variance (Fig. 4). Only $PC_1$ versus $PC_2$ is presented as further PCs did not provide more information. All variables of sensor signal features and tobacco alkaloids were placed between the inner and outer ellipses, denoting 50% and 100% explained variance, respectively, indicating that they were well explained by the model. Marked sensors with small circles show significant variables ($P < 0.05$).

Fig. 4 shows that eleven signal features (IN2~IN6 and K1~K6) marked with small circles were located on the negative factor 1, meanwhile other fifteen signal features (IN7~IN16, K7~K9 and K15) and tobacco alkaloids were located on the positive factor 1. These results indicate that tobacco alkaloids were significantly and negatively correlated to the above eleven signal features, but positively correlated to the other fifteen signal features.

Further investigation of the contribution of sensor signal features to each alkaloid was carried out by PLS1 analysis on jack-knife uncertainty test. The results are reflected in Table. 3, the signal features marked with asterisk indicate the significant features. For instance, ten signal features (IN1~IN5, IN7, IN15, K6, K7 and K15) showed significant sensitivity to nicotine.

### 3.6. Predictability of tobacco alkaloids through electronic nose responses

PLS1 analysis was done to further investigate the predictability of alkaloids using electronic nose responses. The significant $IN_{max}$ and slope (K) were

simultaneously incorporated in the modeling without missing relevant additional information. Eight mathematical models were established based on the calibration/optimization data set, in which, each alkaloid was designated as Y-variable and sensor signal feature responses as predictor X-variable. The established prediction equation was a multiple regression curve (shown in Table 3). For instance, nicotine prediction equation was that:

$Y_{nicotine}$=8.932$IN1_{max}$-24.366$IN2_{max}$-16.375$IN3_{max}$-30.175$IN4_{max}$-26.194$IN5_{max}$-268.0 14$IN6_{max}$+13.838$IN7_{max}$+45.797$IN8_{max}$-9.269$IN9_{max}$+28.703$IN10_{max}$+5.853$IN11_{max}$+7 .517$IN12_{max}$+5.091$IN13_{max}$-52.394$I14_{max}$+43.829$IN15_{max}$-23.547$IN16_{max}$-267.587K1- 88.832K2-2.829K3-34.808K4+26.671K5-1794.000K6+248.306K7-68.883K8-72.012 K9+1981.000 K15-88.670

The predictive performance of these equations was estimated by the parameters of the fitted linear calibration and validated models (Table 4).

For the fitted linear calibration models, the correlation coefficients ($R_{cal}$) represented by the correlation of mean data and regression model, were greater than 0.93 ($R_{cal}\geq 0.93$), while, the regression coefficients of linear calibration models ($R^2_{cal}$) were greater than 0.87 ($R^2_{cal}\geq 0.87$) for tobacco alkaloids (except nicotyrine), which indicating well fit to the calibration model (Table 4). The calibrated parameters ($R_{cal}$ of 0.99, $R^2_{cal}$ of 0.98) for nicotine were very satisfactory, indicating there was a much better fit to nicotine calibration model. However, the $R_{cal}$ of 0.73 and $R^2_{cal}$ of 0.53 for nicotyrine indicate slightly poor fit to the calibration model (Table 4).

For the fitted linear validated models, they were well fitted for nicotine,

myosmine, nornicotine, anabasine, anatabine, 2,3-dipyridyl and cotinine (Table 4), because their correlation coefficients ($R_{val}$) were greater than 0.88 ($R_{val} \geq 0.88$) (Table 4). The regression coefficient of linear validated equation ($R^2_{val}$), used to check the adequacy of the model, represents how successfully the cross-validated regression line approximated raw data points. The value of $R^2_{val}$ for nicotine was 0.94, indicating the established nicotine model has good prediction performance. The $R^2_{val}$ of other alkaloids (except nicotyrine) was greater than 0.80 ($R^2_{val} \geq 0.80$), these results indicate that established models were capable to do prediction for these alkaloids.

Moreover, slight poor prediction capability was shown for nicotyrine due to the relative low $R_{val}$ and $R^2_{val}$ values ($R_{val} = 0.64$ and $R^2_{val} = 0.46$). Possible reason might be that electronic nose sensors were less sensitive towards nicotyrine due to the functional group of analytes. Previous research reported that sensors of electronic nose were less sensitive towards 1-penten-3-ol, hexanoic acid, heptanoic acid, and 2-hexyl-thiophene, etc [22]. These might be an explanation to the relative low $R_{cal}$ and $R^2_{cal}$ of nicotyrine.

### 3.7. Validity checking of established prediction models

The predicted value that gained from prediction models and the reference value (or observed value) that determined by GC-MS analysis were compared to verify the validity of the established models through the analysis of other independent data set. The predicted against reference/observed values were illustrated in Fig. 5.

From Fig. 5A, it was observed that nicotine reference data points were closer to the regression line, which indicate that nicotine reference values and predicted

nicotine values are in good agreement. Root mean square error of prediction (RMSEP) represented the accuracy of prediction model, indicates the average difference between predicted values and reference values [28]. It is worth to mention that a model with large $R^2$ and low RMSEP value is considered to be a good model. Fig. 5A shows that the correlation coefficient (R=0.99), regression coefficient ($R^2$=0.96), and RMSEP (1.25) are quite satisfactory in the validation of nicotine model. These results confirm that established model provides about 96% predictability for nicotine.

For the established models for myosmine, nornicotine, anabasine, 2,3-dipyridyl and cotinine, the validity checking results showed that these models have high correlation coefficient ($R \geq 0.95$) and regression coefficient ($R^2 \geq 0.73$), and low RMSEP ($\leq 0.08$) (Fig. 5). These results indicate that established prediction models were suitable to perform prediction, and they have provided predictability for myosmine of 90% (B), nornicotine of 73% (C), anabasine of 72% (D), 2,3-dipyridyl (G) of 73% and cotinine of 83% (H).

In addition, Fig. $5F_1$ shows that correlation coefficient ($R=0.97$) and RMSEP (about 0.08) for anatabine are considered satisfactory. However, its regression coefficient ($R^2 \geq 0.57$) is relative lower. The finding by analysis is that three samples' reference value points were outliers. The $R^2$ of anatabine (Fig. $5F_2$, $R^2$=0.67) was improved after the outlier samples were removed. Although the improved $R^2$ of anatabine was still not satisfactory, it was acceptable.

Ideally, predicted value should be equal to reference value. Actually, there has been always existed deviation between predicted and measured value. The predicted

value higher or lower than measured value within a certain range is allowed. Hui hong [33] claimed that the average relative deviation between predicted results and reference results of less than 10% is considered acceptable. In present study, relative deviations were all less than 10% for nicotine (A), myosmine (B), nornicotine (C), anabasine (D), anatabine (F), 2,3-dipyridyl (G) and cotinine (H).

## 4. Conclusions

This paper aimed to establish prediction models of tobacco alkaloids by electronic nose system and PLSR analysis for rapid controlling nicotine alkaloids level. Eight alkaloids in tobacco were identified in selected ion scanning module (SIM) and different time segments by gas chromatography-triple quadrupole mass spectrometry (GC-TriQ-MS). The content of which was found significantly varied ($P < 0.05$) with leaf positions.

Three signal features of electronic nose sensors (maximum intensity ($IN_{max}$), slope (K) and the time of the maximum intensity occurred ($T_{max}$) were extracted for prediction assessment of tobacco alkaloids. The significant features were used in PLSR analysis to establish prediction model for improving the predictive capability of established models without losing relevant additional information. Prediction models were established for predicting tobacco alkaloids level, and satisfying results were obtained for nicotine, myosmine, nornicotine, anabasine, anatabine, 2,3-dipyridyl and cotinine.

In addition, other independent data set was employed to check the validity of established models, and good predictability for nicotine, myosmine, nornicotine,

anabasine, anatabine, 2,3-dipyridyl and cotinine were confirmed.

Present study demonstrated that Fox 4000 electronic nose is capable of analyzing the alkaloids level in tobacco without laborious sample pretreatment. However, further study is still needed to improve the precision and robustness of the alkaloids calibration models. This work provided evidence that electronic nose could be used as a testing tool to rapidly and quantitatively detect nicotine alkaloids content in tobacco.

**Conflict of Interest**

The authors declare that there are no conflicts of interest.

## References


1. J. G. Lisko, S. B. Stanfill, B. W. Duncan and C. H. Watson, *Anal. Chem.*, 2013, 85, 3380-3384.
2. R. S. Lewis, R. G. Parker, D. A. Danehower, K. Andres, A. M. Jack, D. S. Whitley and L. P. Bush, *J. Agric. Food Chem.*, 2012, 60, 6454-6461.
3. K. J. Clemens, S. Caille, L. Stinus and M. Cador, *Int. J. Neuropsychopharmacol.*, 2009, 12, 1355-1366.
4. D. K. Hatsukami, N. L. Benowitz, E. Donny, J. Henningfield and M. Zeller, *Nicotine & Tobacco Research*, 2013, 15, 1003-1013.
5. B. Xie, P. Palmer, Y. Li, C. Lin and C. A. Johnson, *Nicotine Tob Res*, 2013, 15, 1673-1681.
6. J. Zhang, S. Sokhansanj, S. Wu, R. Fang and W. Yang, *Computers and Electronics in Agriculture*, 1997, 16, 231-244.
7. R. Jeffrey and T. Tso, *Journal of Agricultural and Food Chemistry*, 1955, 3, 680-682.
8. W. Sun, Z. Zhou, Y. Li, Z. Xu, W. Xia and F. Zhong, *Eur. Food Res. Technol.*, 2012, 235, 745-752.
9. K. Cai, Z. Xiang, J. Zhang, S. Zhou, Y. Feng and Z. Geng, *Analytical Methods*, 2012, 4, 2095-2100.
10. L. Fischer, F. Mikus, R. Jantos and G. Skopp, *International Journal of Legal Medicine*, 2014, 129, 279-287.
11. S. Deshmukh, R. Bandyopadhyay, N. Bhattacharyya, R. A. Pandey and A. Jana, *Talanta*, 2015, 144, 329-340.
12. S. D. Rodríguez, D. A. Barletta, T. F. Wilderjans and D. L. Bernik, *Food Analytical Methods*, 2014, 7, 2042-2050.
13. Y. Xiao, J. Jiaojiao, H. Guohua, Y. Fangyuan, W. Minmin, H. Jie, Y. Xiaoguo and D. Shanggui, *Food Analytical Methods*, 2014, 7, 1612-1618.
14. S. Xu, Z. Zhou, H. Lu, X. Luo and Y. Lan, *Sensors*, 2014, 14, 5486-5501.
15. Z. Xiao, D. Yu, Y. Niu, F. Chen, S. Song, J. Zhu and G. Zhu, *Journal of Chromatography B*, 2014, 945-946, 92-100.
16. K. Brudzewski, S. Osowski and A. Golembiecka, *Expert Syst.Appl.*, 2012, 39, 9886-9891.
17. C. Di Natale, A. Macagnano, E. Martinelli, R. Paolesse, E. Proietti and A. D'Amico, *Sensors and Actuators B: Chemical*, 2001, 78, 26-31.
18. P. Ködderitzsch, R. Bischoff, P. Veitenhansl, W. Lorenz and G. Bischoff, *Sensor and Actuat B-Chem.*, 2005, 107, 479-489.
19. R. Feldhoff, C. A. Saby and P. Bernadet, *Flavour and fragrance journal*, 2000, 15, 215-222.
20. A. Kirsching, G. Bazar, S. Szvath and R. Romvari, *Agriculturae Conspectus Scientificus (ACS)*, 2011, 76, 311-315.
21. M. Su, B. Zhang, Z. Ye, K. Chen, J. Guo, X. Gu and J. Shen, *Scientia Horticulturae*, 2013, 150, 146-153.
22. S. Song, X. Zhang, K. Hayat, C. Jia, S. Xia, F. Zhong, Z. Xiao, H. Tian and Y. Niu, *Sensor and Actuat B-Chem.*, 2010, 147, 660-668.
23. S. Song, Q. Tang, K. Hayat, E. Karangwa, X. Zhang and Z. Xiao, *Meat Sci*, 2014, 96, 1191-1200.
24. J. Qin, L. Xie and Y. Ying, *Food Chemistry*, 2015, 170, 415-422.
25. S. Poutiainen, S. Matero, T. Hämäläinen, J. Leskinen, J. Ketolainen and K. Järvinen, *Powder*

*Technology*, 2012, 228, 149-157.

26. J. C. M. Pires, M. C. M. Alvim-Ferraz, M. C. Pereira and F. G. Martins, *Journal of Statistical Computation and Simulation*, 2012, 82, 183-192.

27. K. Cai, Z. Xiang, J. Zhang, S. Zhou, Y. Feng and Z. Geng, *Anal. Methods*, 2012, 4, 2095.

28. A. Z. Berna, S. Trowell, W. Cynar and D. Cozzolino, *J. Agric. Food Chem.*, 2008, 56, 3238-3244.

29. Z. Zhou, Z. Xu, J. Shu, W. Sun, C. Yin, M. Chen, Y. Li and F. Zhong, *Anal. Methods*, 2013, 5, 3557-3564.

30. R. Marchetti, F. Castelli and R. Contillo, *Agronomy Journal*, 2006, 98, 666-674.

31. R. Feldhoff, P. Bernadet and C.-A. Saby, *Analyst*, 1999, 124, 1167-1173.

32. S. Deshmukh, A. Jana, N. Bhattacharyya, R. Bandyopadhyay and R. A. Pandey, *Atmospheric Environment*, 2014, 82, 401-409.

33. H. Hong, Y. Luo, S. Zhu and H. Shen, *International Journal of Food Science & Technology*, 2012, 47, 488-494.

Table 1. Qualitative and quantitative analysis of eight alkaloids.

| NO. | Alkaloids | Time | ID [a] | RI [b] | MRI [c] | Scan segment (min) | Quantitative ion (SIM, m/z) | Range (mg. g$^{-1}$) | (mg. g$^{-1}$) [d] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | BF | CF | XF |
| 1 | Nicotine | 7.33 | A | 1360 | 1360 | 7.00～8.50 | 163,84,133 | 41.65-63.67 | 61.73±4.35$^{c}$ | 50.84±2.12$^{b}$ | 39.98±2.80$^{a}$ |
| 2 | Myosmine | 8.74 | B | 1427 | 1430 | 8.50～9.25 | 159,118,78 | 0.02-0.06 | 0.05±0.01$^{b}$ | 0.04±0.01$^{ab}$ | 0.03±0.01$^{a}$ |
| 3 | Nornicotine | 9.13 | A | 1435.4 | 1435 | 9.25～10.00 | 147,119,70 | 0.31-0.84 | 0.74±0.05$^{c}$ | 0.57±0.02$^{b}$ | 0.46±0.07$^{a}$ |
| 4 | Anabasine | 10.81 | A | 1525 | 1527 | 10.00～11.50 | 84,106,133 | 0.15-0.31 | 0.26±0.02$^{b}$ | 0.22±0.04$^{a}$ | 0.21±0.03$^{a}$ |
| 5 | Nicotyrine | 10.92 | B | 1488 | 1490 | 10.00～11.50 | 158,130,116 | 0.05-0.12 | 0.09±0.02$^{b}$ | 0.08±0.01$^{b}$ | 0.06±0.01$^{a}$ |
| 6 | Anatabine | 11.61 | A | -- | 1510 | 11.50～12.25 | 131,106,160 | 0.58-1.25 | 1.05±0.09$^{b}$ | 0.90±0.07$^{a}$ | 0.83±0.05$^{a}$ |
| 7 | 2,3-Dipyridyl | 12.23 | B | 1536 | 1540 | 12.25～13.00 | 156,130 | 0.01-0.04 | 0.03±0.00$^{b}$ | 0.03±0.01$^{b}$ | 0.02±0.00$^{a}$ |
| 8 | Cotinine | 14.20 | C | -- | 1605 | 13.00～15.00 | 147,133,121 | 0.02-0.04 | 0.03±0.00$^{b}$ | 0.03±0.00$^{b}$ | 0.02±0.00$^{a}$ |

[a] The identification is indicated by the following symbols: (A) mass spectrum and RI agree with that of the authentic standard compositions run under similar GC-MS conditions; (B) mass spectrum and RI agree with NIST Standard Reference Database (http://webbook.nist.gov/chemistry/); (C) tentative identification based on interpretation of mass spectrum.

[b] RI, Kovata index reference from NIST Standard Reference Database, that the compositions were determined on non-polar column (HP/DB-5)column run under similar GC-MS conditions.

[c] MRI, Kovata index were determined by using a series hydrocarbons of C8～C40 on the DB-5MS column described on Section 2.2.

[d] Approximate concentrations (mean ± standard deviation, average of triplicate) for each alkaloid, different letters within a row denote significantly different at $P<0.05$ level.

Table 2 Sensor signal feature values of maximum intensity ($IN_{max}$) and slope (K) for flue-cured tobacco samples

| | $IN1_{max}$ | $IN2_{max}$ | $IN3_{max}$ | $IN4_{max}$ | $IN5_{max}$ | $IN6_{max}$ | $IN7_{max}$ | $IN8_{max}$ | $IN9_{max}$ | $IN10_{max}$ | $IN11_{max}$ | $IN12_{max}$ | $IN13_{max}$ | $IN14_{max}$ | $IN15_{max}$ | $IN16_{max}$ | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | K15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| B1 | 0.066 | -0.440 | -0.519 | -0.410 | -0.448 | -0.041 | 0.531 | 0.501 | 0.466 | 0.557 | 0.435 | 0.582 | 0.585 | 0.301 | 0.492 | 0.200 | 0.002 | -0.073 | -0.086 | -0.051 | -0.064 | -0.005 | 0.031 | 0.033 | 0.027 | 0.014 |
| B2 | 0.062 | -0.408 | -0.495 | -0.387 | -0.412 | -0.040 | 0.542 | 0.504 | 0.474 | 0.561 | 0.448 | 0.593 | 0.591 | 0.301 | 0.522 | 0.203 | 0.002 | -0.051 | -0.083 | -0.039 | -0.059 | -0.004 | 0.032 | 0.034 | 0.030 | 0.014 |
| B3 | 0.066 | -0.422 | -0.506 | -0.391 | -0.422 | -0.041 | 0.527 | 0.503 | 0.473 | 0.559 | 0.436 | 0.581 | 0.585 | 0.305 | 0.499 | 0.203 | 0.002 | -0.056 | -0.078 | -0.039 | -0.070 | -0.005 | 0.031 | 0.037 | 0.031 | 0.014 |
| B4 | 0.060 | -0.400 | -0.487 | -0.380 | -0.405 | -0.040 | 0.542 | 0.506 | 0.475 | 0.561 | 0.449 | 0.595 | 0.589 | 0.300 | 0.522 | 0.200 | 0.002 | -0.057 | -0.097 | -0.042 | -0.068 | -0.005 | 0.034 | 0.036 | 0.030 | 0.014 |
| B5 | 0.060 | -0.409 | -0.494 | -0.388 | -0.413 | -0.038 | 0.540 | 0.502 | 0.471 | 0.559 | 0.445 | 0.590 | 0.591 | 0.302 | 0.520 | 0.203 | 0.002 | -0.058 | -0.082 | -0.039 | -0.069 | -0.005 | 0.032 | 0.039 | 0.031 | 0.015 |
| B6 | 0.065 | -0.424 | -0.514 | -0.400 | -0.427 | -0.039 | 0.559 | 0.515 | 0.489 | 0.569 | 0.469 | 0.613 | 0.605 | 0.311 | 0.495 | 0.211 | 0.002 | -0.061 | -0.083 | -0.041 | -0.075 | -0.005 | 0.035 | 0.040 | 0.033 | 0.013 |
| B7 | 0.067 | -0.419 | -0.503 | -0.395 | -0.427 | -0.040 | 0.524 | 0.499 | 0.467 | 0.555 | 0.430 | 0.578 | 0.576 | 0.298 | 0.492 | 0.197 | 0.002 | -0.060 | -0.084 | -0.040 | -0.071 | -0.004 | 0.029 | 0.036 | 0.029 | 0.013 |
| B8 | 0.062 | -0.430 | -0.516 | -0.403 | -0.433 | -0.042 | 0.556 | 0.514 | 0.484 | 0.569 | 0.464 | 0.607 | 0.606 | 0.311 | 0.496 | 0.210 | 0.002 | -0.054 | -0.103 | -0.045 | -0.072 | -0.004 | 0.029 | 0.037 | 0.030 | 0.012 |
| C1 | 0.070 | -0.423 | -0.502 | -0.396 | -0.429 | -0.040 | 0.510 | 0.492 | 0.463 | 0.549 | 0.417 | 0.565 | 0.570 | 0.299 | 0.475 | 0.198 | 0.003 | -0.053 | -0.100 | -0.044 | -0.071 | -0.004 | 0.027 | 0.035 | 0.029 | 0.012 |
| C2 | 0.067 | -0.411 | -0.490 | -0.385 | -0.416 | -0.040 | 0.507 | 0.496 | 0.469 | 0.545 | 0.417 | 0.564 | 0.558 | 0.297 | 0.472 | 0.198 | 0.003 | -0.059 | -0.082 | -0.043 | -0.069 | -0.004 | 0.027 | 0.035 | 0.031 | 0.012 |
| C3 | 0.070 | -0.405 | -0.485 | -0.380 | -0.412 | -0.039 | 0.498 | 0.492 | 0.463 | 0.549 | 0.408 | 0.556 | 0.563 | 0.299 | 0.469 | 0.196 | 0.003 | -0.045 | -0.081 | -0.038 | -0.059 | -0.004 | 0.026 | 0.031 | 0.026 | 0.013 |
| C4 | 0.063 | -0.406 | -0.485 | -0.382 | -0.412 | -0.039 | 0.512 | 0.494 | 0.463 | 0.551 | 0.419 | 0.569 | 0.567 | 0.294 | 0.478 | 0.193 | 0.002 | -0.058 | -0.097 | -0.048 | -0.069 | -0.004 | 0.028 | 0.033 | 0.029 | 0.013 |
| C5 | 0.062 | -0.384 | -0.465 | -0.366 | -0.390 | -0.038 | 0.496 | 0.490 | 0.458 | 0.547 | 0.404 | 0.556 | 0.550 | 0.288 | 0.461 | 0.189 | 0.002 | -0.048 | -0.093 | -0.046 | -0.065 | -0.004 | 0.028 | 0.035 | 0.027 | 0.012 |
| C6 | 0.070 | -0.408 | -0.488 | -0.383 | -0.412 | -0.039 | 0.512 | 0.493 | 0.469 | 0.549 | 0.421 | 0.567 | 0.580 | 0.308 | 0.490 | 0.203 | 0.002 | -0.058 | -0.081 | -0.043 | -0.069 | -0.004 | 0.027 | 0.033 | 0.028 | 0.013 |
| C7 | 0.064 | -0.411 | -0.491 | -0.385 | -0.417 | -0.039 | 0.506 | 0.492 | 0.465 | 0.548 | 0.415 | 0.565 | 0.559 | 0.294 | 0.471 | 0.193 | 0.003 | -0.051 | -0.082 | -0.039 | -0.069 | -0.004 | 0.028 | 0.033 | 0.026 | 0.012 |
| C8 | 0.069 | -0.395 | -0.433 | -0.371 | -0.401 | -0.039 | 0.491 | 0.488 | 0.458 | 0.545 | 0.401 | 0.548 | 0.574 | 0.295 | 0.452 | 0.192 | 0.003 | -0.049 | -0.080 | -0.038 | -0.060 | -0.004 | 0.026 | 0.030 | 0.025 | 0.012 |
| X1 | 0.060 | -0.387 | -0.468 | -0.365 | -0.391 | -0.038 | 0.469 | 0.489 | 0.464 | 0.546 | 0.380 | 0.526 | 0.525 | 0.288 | 0.458 | 0.184 | 0.003 | -0.048 | -0.078 | -0.037 | -0.065 | -0.003 | 0.026 | 0.033 | 0.026 | 0.012 |
| X2 | 0.059 | -0.384 | -0.464 | -0.364 | -0.389 | -0.037 | 0.469 | 0.487 | 0.461 | 0.545 | 0.378 | 0.525 | 0.525 | 0.286 | 0.456 | 0.182 | 0.003 | -0.048 | -0.077 | -0.036 | -0.065 | -0.003 | 0.026 | 0.032 | 0.026 | 0.012 |
| X3 | 0.065 | -0.388 | -0.466 | -0.368 | -0.394 | -0.037 | 0.472 | 0.489 | 0.457 | 0.548 | 0.380 | 0.527 | 0.524 | 0.280 | 0.448 | 0.178 | 0.003 | -0.043 | -0.078 | -0.037 | -0.056 | -0.004 | 0.025 | 0.031 | 0.025 | 0.012 |
| X4 | 0.049 | -0.391 | -0.470 | -0.368 | -0.396 | -0.039 | 0.465 | 0.485 | 0.459 | 0.543 | 0.374 | 0.520 | 0.521 | 0.286 | 0.450 | 0.182 | 0.003 | -0.043 | -0.078 | -0.037 | -0.057 | -0.004 | 0.024 | 0.032 | 0.027 | 0.012 |
| X5 | 0.055 | -0.344 | -0.423 | -0.331 | -0.348 | -0.034 | 0.448 | 0.478 | 0.446 | 0.537 | 0.379 | 0.527 | 0.528 | 0.283 | 0.435 | 0.180 | 0.002 | -0.043 | -0.060 | -0.033 | -0.058 | -0.003 | 0.026 | 0.034 | 0.028 | 0.012 |
| X6 | 0.057 | -0.343 | -0.421 | -0.330 | -0.347 | -0.035 | 0.444 | 0.482 | 0.451 | 0.540 | 0.376 | 0.524 | 0.522 | 0.282 | 0.429 | 0.180 | 0.002 | -0.043 | -0.070 | -0.033 | -0.050 | -0.003 | 0.028 | 0.032 | 0.027 | 0.011 |
| X7 | 0.060 | -0.348 | -0.431 | -0.334 | -0.351 | -0.040 | 0.461 | 0.476 | 0.454 | 0.547 | 0.392 | 0.537 | 0.554 | 0.298 | 0.432 | 0.194 | 0.002 | -0.040 | -0.072 | -0.039 | -0.050 | -0.003 | 0.026 | 0.032 | 0.027 | 0.011 |

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X8 | 0.055 | -0.342 | -0.415 | -0.325 | -0.345 | -0.034 | 0.457 | 0.476 | 0.450 | 0.533 | 0.375 | 0.523 | 0.527 | 0.287 | 0.425 | 0.182 | 0.002 | -0.043 | -0.069 | -0.033 | -0.049 | -0.003 | 0.025 | 0.032 | 0.026 | 0.011 |
| TB1 | 0.066 | -0.428 | -0.502 | -0.395 | -0.422 | -0.040 | 0.527 | 0.498 | 0.471 | 0.555 | 0.434 | 0.579 | 0.586 | 0.305 | 0.505 | 0.203 | 0.003 | -0.045 | -0.082 | -0.039 | -0.059 | -0.004 | 0.028 | 0.031 | 0.026 | 0.014 |
| TB2 | 0.066 | -0.420 | -0.501 | -0.395 | -0.425 | -0.040 | 0.518 | 0.497 | 0.468 | 0.553 | 0.447 | 0.593 | 0.568 | 0.302 | 0.489 | 0.199 | 0.002 | -0.051 | -0.098 | -0.048 | -0.069 | -0.004 | 0.029 | 0.036 | 0.028 | 0.013 |
| TB3 | 0.067 | -0.430 | -0.518 | -0.399 | -0.433 | -0.041 | 0.546 | 0.505 | 0.479 | 0.560 | 0.473 | 0.606 | 0.599 | 0.309 | 0.497 | 0.198 | 0.003 | -0.046 | -0.082 | -0.039 | -0.060 | -0.004 | 0.027 | 0.036 | 0.026 | 0.014 |
| TB4 | 0.065 | -0.419 | -0.501 | -0.386 | -0.415 | -0.039 | 0.518 | 0.498 | 0.468 | 0.580 | 0.465 | 0.616 | 0.609 | 0.305 | 0.505 | 0.198 | 0.003 | -0.051 | -0.082 | -0.039 | -0.069 | -0.004 | 0.029 | 0.033 | 0.026 | 0.014 |
| TB5 | 0.064 | -0.409 | -0.494 | -0.387 | -0.407 | -0.038 | 0.504 | 0.494 | 0.466 | 0.557 | 0.459 | 0.625 | 0.590 | 0.296 | 0.504 | 0.194 | 0.002 | -0.049 | -0.068 | -0.037 | -0.066 | -0.004 | 0.030 | 0.035 | 0.029 | 0.013 |
| TB6 | 0.063 | -0.406 | -0.488 | -0.375 | -0.402 | -0.039 | 0.502 | 0.490 | 0.457 | 0.547 | 0.468 | 0.604 | 0.566 | 0.303 | 0.500 | 0.192 | 0.002 | -0.057 | -0.096 | -0.047 | -0.067 | -0.004 | 0.028 | 0.033 | 0.029 | 0.014 |
| TC1 | 0.061 | -0.415 | -0.508 | -0.397 | -0.429 | -0.040 | 0.514 | 0.490 | 0.461 | 0.547 | 0.423 | 0.579 | 0.561 | 0.295 | 0.475 | 0.194 | 0.002 | -0.048 | -0.078 | -0.037 | -0.056 | -0.004 | 0.028 | 0.033 | 0.027 | 0.012 |
| TC2 | 0.062 | -0.424 | -0.510 | -0.391 | -0.420 | -0.039 | 0.506 | 0.486 | 0.460 | 0.544 | 0.417 | 0.571 | 0.549 | 0.287 | 0.457 | 0.187 | 0.003 | -0.050 | -0.080 | -0.038 | -0.068 | -0.004 | 0.028 | 0.032 | 0.025 | 0.012 |
| TC3 | 0.066 | -0.425 | -0.505 | -0.390 | -0.417 | -0.039 | 0.502 | 0.488 | 0.459 | 0.545 | 0.402 | 0.559 | 0.555 | 0.295 | 0.460 | 0.192 | 0.002 | -0.048 | -0.077 | -0.036 | -0.056 | -0.004 | 0.031 | 0.033 | 0.027 | 0.012 |
| TC4 | 0.063 | -0.407 | -0.494 | -0.375 | -0.401 | -0.040 | 0.502 | 0.486 | 0.468 | 0.544 | 0.411 | 0.567 | 0.551 | 0.294 | 0.456 | 0.191 | 0.002 | -0.055 | -0.077 | -0.037 | -0.056 | -0.004 | 0.027 | 0.035 | 0.029 | 0.012 |
| TC5 | 0.061 | -0.408 | -0.496 | -0.379 | -0.393 | -0.039 | 0.503 | 0.488 | 0.459 | 0.545 | 0.413 | 0.565 | 0.552 | 0.292 | 0.463 | 0.189 | 0.002 | -0.054 | -0.076 | -0.036 | -0.064 | -0.004 | 0.029 | 0.033 | 0.029 | 0.013 |
| TC6 | 0.063 | -0.401 | -0.487 | -0.373 | -0.395 | -0.039 | 0.513 | 0.491 | 0.463 | 0.548 | 0.413 | 0.569 | 0.558 | 0.294 | 0.472 | 0.193 | 0.002 | -0.043 | -0.079 | -0.037 | -0.049 | -0.004 | 0.028 | 0.033 | 0.027 | 0.012 |
| TX1 | 0.065 | -0.386 | -0.467 | -0.336 | -0.392 | -0.038 | 0.493 | 0.489 | 0.459 | 0.546 | 0.403 | 0.550 | 0.553 | 0.293 | 0.459 | 0.191 | 0.002 | -0.043 | -0.078 | -0.041 | -0.056 | -0.004 | 0.026 | 0.033 | 0.029 | 0.012 |
| TX2 | 0.060 | -0.380 | -0.463 | -0.346 | -0.387 | -0.037 | 0.489 | 0.484 | 0.453 | 0.543 | 0.395 | 0.546 | 0.541 | 0.283 | 0.454 | 0.183 | 0.003 | -0.042 | -0.077 | -0.036 | -0.055 | -0.004 | 0.026 | 0.030 | 0.025 | 0.012 |
| TX3 | 0.065 | -0.388 | -0.460 | -0.347 | -0.388 | -0.036 | 0.492 | 0.488 | 0.461 | 0.545 | 0.387 | 0.529 | 0.543 | 0.295 | 0.460 | 0.192 | 0.003 | -0.049 | -0.078 | -0.037 | -0.065 | -0.003 | 0.027 | 0.033 | 0.026 | 0.012 |
| TX4 | 0.066 | -0.387 | -0.416 | -0.325 | -0.352 | -0.038 | 0.490 | 0.488 | 0.460 | 0.545 | 0.400 | 0.547 | 0.554 | 0.296 | 0.457 | 0.193 | 0.002 | -0.043 | -0.078 | -0.037 | -0.056 | -0.003 | 0.026 | 0.033 | 0.027 | 0.012 |
| TX5 | 0.064 | -0.344 | -0.424 | -0.334 | -0.378 | -0.037 | 0.483 | 0.484 | 0.458 | 0.541 | 0.395 | 0.542 | 0.550 | 0.296 | 0.454 | 0.192 | 0.002 | -0.053 | -0.075 | -0.035 | -0.063 | -0.004 | 0.027 | 0.035 | 0.027 | 0.011 |
| TX6 | 0.062 | -0.370 | -0.420 | -0.342 | -0.374 | -0.037 | 0.482 | 0.482 | 0.455 | 0.540 | 0.393 | 0.541 | 0.548 | 0.293 | 0.453 | 0.190 | 0.002 | -0.046 | -0.090 | -0.035 | -0.062 | -0.004 | 0.027 | 0.034 | 0.027 | 0.012 |

In the table, $IN1_{max}$~$IN16_{max}$ indicate the maximum intensity of sensor 1 ~ sensor 16. K1 ~ K9 and K15 indicate the slope of sensor 1 ~ sensor 9 and sensor 15, respectively. The numbered sensors denote (1): LY2/LG, (2): LY2/G, (3): LY2/AA, (4): LY2/GH, (5): LY2/gCTl, (6): LY2/gCT, (7): T30/1, (8): P10/1, (9): P10/2, (10): P40/1, (11): T70/2, (12): PA/2, (13): P30/1, (14): P40/2, (15): P30/2 and (16): T40/2 in this paper. Total forty-two samples sourced from fourteen different origins and three leaf positions (upper B, middle C, and lower X leaves) were used in this study. These samples were divided into two groups, the first group contained twenty-four samples, named as: B1～8, C1～8 and X1～8, that sourced from eight origins, followed as 1: Shaoyang, 2: Longhui, 3: Chenzhou, 4: Xingyi, 5: Zhengan, 6: Changning, 7: Wenshan and 8: Fengjie city, and three leaf positions (B, C and X) respectively, which were used for tobacco alkaloids' characterization analysis and model calibration/optimization. The second group contained eighteen samples, named as TB1～6, TC1～6 and TX1～6, that originated from six origins, followed as 1:Yongxing, 2: Panxian, 3: Zunyi, 4: Baoshan, 5: Anning and 6: Yuqing City and three leaf positions (B, C and X) respectively, which were used for models' validation.

Table 3. Prediction equations of tobacco alkaloids based on electronic nose sensor responses by $PLS_1$ cross-validation analysis.

| Signal features | Tobacco alkaloids (Y-Variables) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| (X-Variables) | Nicotine | Myosmine | Nornicotine | Anabasine | Nicotyrine | Anatabine | 2,3-Dipyridyl | Cotinine |
| $IN1_{max}$ | 8.932* | -0.048 | 1.332 | 0.430* | 0.071 | 2.051* | 0.108* | -0.105 |
| $IN2_{max}$ | -24.366* | -0.041* | -0.552 | -0.074* | -0.019 | -0.472* | -0.024* | -0.010 |
| $IN3_{max}$ | -16.375* | -0.061 | -0.430 | -0.057* | -0.019 | -0.361* | -0.014* | -0.011* |
| $IN4_{max}$ | -30.175* | -0.057* | -0.662 | -0.086* | -0.022 | -0.537* | -0.026* | -0.014 |
| $IN5_{max}$ | -26.194* | -0.046* | -0.570 | -0.074* | -0.019 | -0.471* | -0.024* | -0.011 |
| $IN6_{max}$ | -268.014 | 0.054 | -4.166 | -1.204* | -0.239 | -4.934* | -0.258* | 0.177 |
| $IN7_{max}$ | 13.838* | 0.019 | 0.230 | 0.017 | 0.016 | 0.118* | 0.007* | 0.005 |
| $IN8_{max}$ | 45.797 | 0.044 | 0.687 | 0.046 | 0.049 | 0.539* | 0.030* | 0.025 |
| $IN9_{max}$ | -9.269 | -0.027 | 0.382 | 0.074 | 0.045 | 0.601* | 0.034* | -0.014 |
| $IN10_{max}$ | 28.703 | 0.107 | 0.976 | 0.112 | 0.055 | 0.640* | 0.028* | 0.001 |
| $IN11_{max}$ | 5.853 | 0.013 | 0.117 | 0.005 | 0.019 | 0.044 | 0.005* | 0.000 |
| $IN12_{max}$ | 7.517 | 0.018 | 0.148 | 0.005 | 0.020 | 0.052 | 0.005 | 0.000 |
| $IN13_{max}$ | 5.091 | 0.000 | 0.180 | 0.021 | 0.019 | 0.072 | 0.008 | -0.006 |
| $IN14_{max}$ | -52.394 | -0.076 | 0.057 | 0.104 | 0.061 | 0.332 | 0.029 | -0.061 |
| $IN15_{max}$ | 43.829* | 0.060 | 0.533 | 0.043* | 0.016 | 0.165* | 0.000 | 0.034* |
| $IN16_{max}$ | -23.547 | -0.057 | 0.001 | 0.080 | 0.059 | 0.236* | 0.022 | -0.046 |
| K1 | -267.587 | -1.423 | 29.483 | 4.988 | -0.385 | 33.695* | 1.699* | -0.831 |
| K2 | -88.832 | -0.219 | -1.010 | -0.175* | -0.076 | -0.588* | -0.023 | -0.089* |
| K3 | -2.829 | 0.107 | -0.805 | -0.116 | -0.029 | -0.477 | -0.043* | 0.069 |
| K4 | -34.808 | 0.072 | -0.848 | -0.467* | -0.105 | -1.434* | -0.096* | 0.139 |
| K5 | 26.671 | 0.039 | -0.841 | -0.213 | -0.051 | -1.411* | -0.062* | 0.007 |
| K6 | -1794.000* | 0.385 | -19.484 | -1.504 | -0.861 | -9.902 | -0.228 | -1.081 |
| K7 | 248.306* | 0.608* | 1.702 | -0.148 | 0.181 | -0.746 | -0.060 | 0.168 |
| K8 | -68.883 | -0.421 | -4.533 | -0.250 | 0.183 | -1.378 | -0.108 | -0.060 |
| K9 | -72.012 | -0.611 | -6.622 | -0.449 | 0.205 | -2.810* | -0.165* | 0.113 |
| K15 | 1981.000* | 2.418* | 18.724 | 1.318 | 0.426 | 4.675 | -0.215 | 1.598* |
| BO | -88.670 | -0.133 | -2.516 | -0.232 | -0.160 | -1.638 | -0.102 | 0.007 |

Prediction equations described as: $Y_{alkaloids}=a_1IN1_{max}+a_2IN2_{max}+a_3IN3_{max}+.....a_{16}IN16_{max}+b_1K1+b_2K2+.....+b_9K9+b_{15}K15+BO$, in which, $IN1_{max}$~$IN16_{max}$ indicate the maximum intensity of sensor 1 ~ sensor 16. K1 ~ K9 and K15 indicate the slope of sensor 1 ~ sensor 9 and sensor 15, respectively. And $a_1$-$a_{16}$, $a_1$-$a_9$ and $a_{15}$ denote the corresponding features coefficient. The value marked with "*" denote the corresponding feature was significant at $P<0.05$ level.

Table 4. Predictive performance of developed equations.

| Alkaloids | Statistical parameter | | | | |
|---|---|---|---|---|---|
| | [a] $R_{cal}$ | [b] $R_{val}$ | [c] RMSEP | [d] $R^2_{cal}$ | [e] $R^2_{val}$ |
| Nicotine | 0.99 | 0.97 | 2.10 | 0.98 | 0.94 |
| Myosmine | 0.96 | 0.89 | 0.005 | 0.93 | 0.82 |
| Nornicotine | 0.95 | 0.91 | 0.066 | 0.90 | 0.84 |
| Anabasine | 0.93 | 0.88 | 0.011 | 0.87 | 0.81 |
| Nicotyrine | 0.73 | 0.64 | 0.011 | 0.53 | 0.46 |
| Anatabine | 0.96 | 0.93 | 0.047 | 0.91 | 0.87 |
| 2,3-Dipyridyl | 0.95 | 0.92 | 0.002 | 0.91 | 0.86 |
| Cotinine | 0.96 | 0.88 | 0.002 | 0. 92 | 0.80 |

[a] $R_{cal}$, denote the correlation coefficients of the data fit with calibration model.

[b] $R_{val}$, denote the correlation coefficients of the data fit with validation model.

[c] RMSEP, root mean square error of prediction.

[d] $R^2_{cal}$ is the raw regression coefficients ($R^2$) of the calibration model.

[e] $R^2_{val}$ is the adjusted regression coefficients ($R^2$) of the validation model.

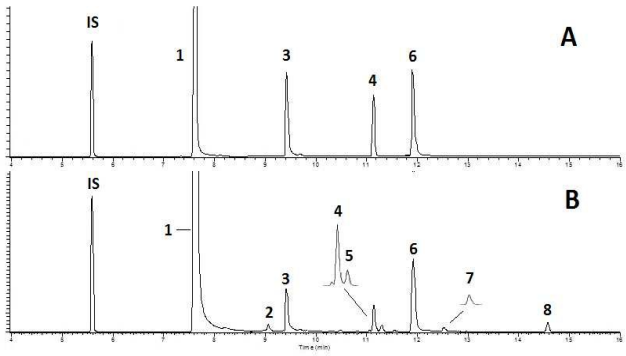



Fig.1 Total ion chromatogram (TIC) of the eight alkaloids. A is the TIC of authentic standard compositions; B is the TIC of tobacco sample in SIM scan mode.

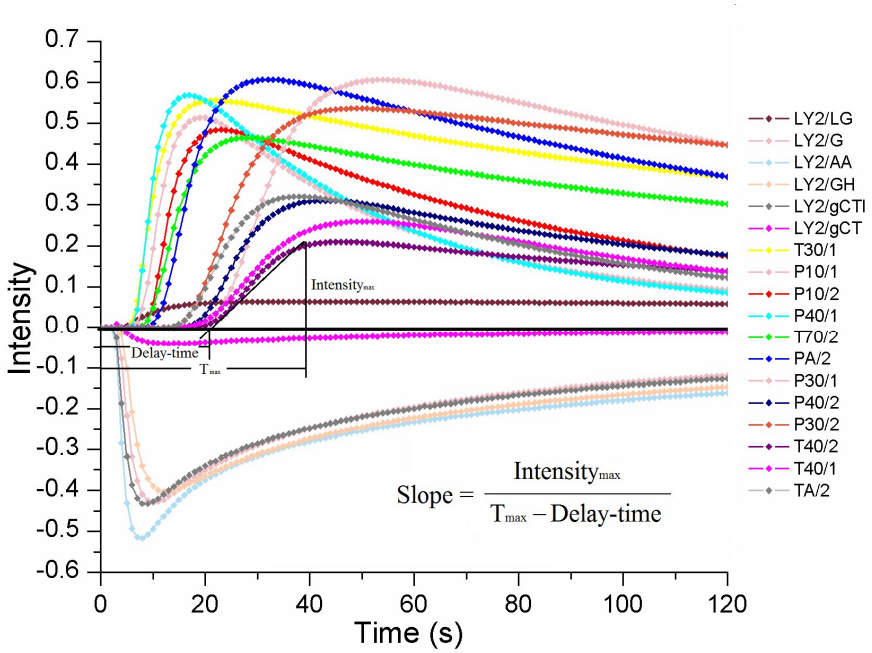



Fig.2 Typical response curves of eighteen sensors of electronic nose for Flue-cured tobacco sample

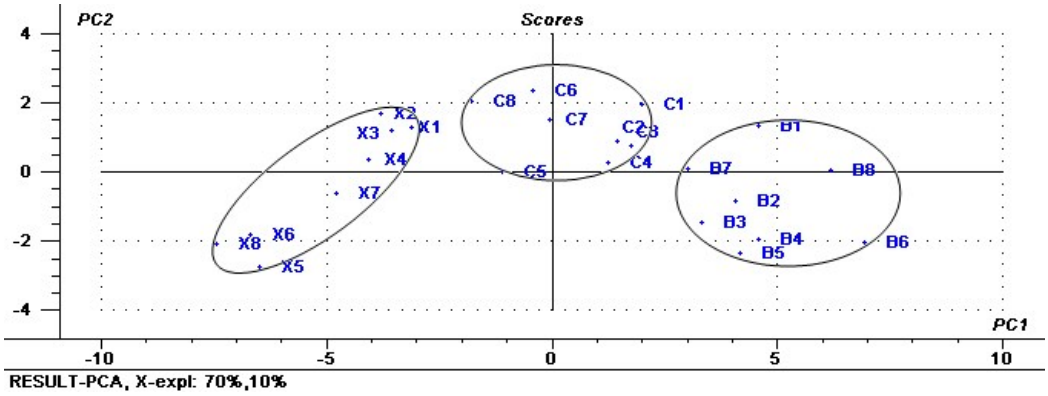



Fig.3 Score plot of PCA analysis for different leaf position samples based on electronic nose analysis. In the figure, B1～8, C1～8 and X1～8 denote tobacco samples that sourced from eight origins (1: Shaoyang, 2: Longhui, 3: Chenzhou, 4: Xingyi, 5: Zhengan, 6: Changning, 7: Wenshan and 8: Fengjie city) and three leaf positions (B, C and X), respectively.

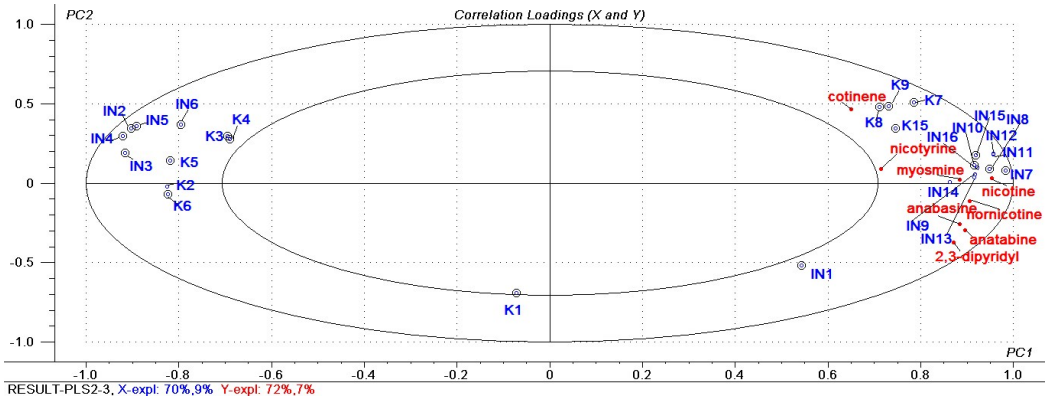



Fig.4 An overview of the variables in PLSR correlation loadings plot. The signal features of electronic nose sensors were designed as X-matrix, eight alkaloids were designed as Y-matrix. In the figure, $IN1_{max}$~$IN16_{max}$ indicate the maximum intensity of sensor 1 ~ sensor 16. K1~K9 and K15 indicate the slope of sensor 1 ~ sensor 9 and sensor 15, respectively. The numbered sensors denote (1): LY2/LG, (2): LY2/G, (3): LY2/AA, (4): LY2/GH, (5): LY2/gCTl, (6): LY2/gCT, (7): T30/1, (8): P10/1, (9): P10/2, (10): P40/1, (11): T70/2, (12): PA/2, (13): P30/1, (14): P40/2, (15): P30/2 and (16): T40/2 in this paper.

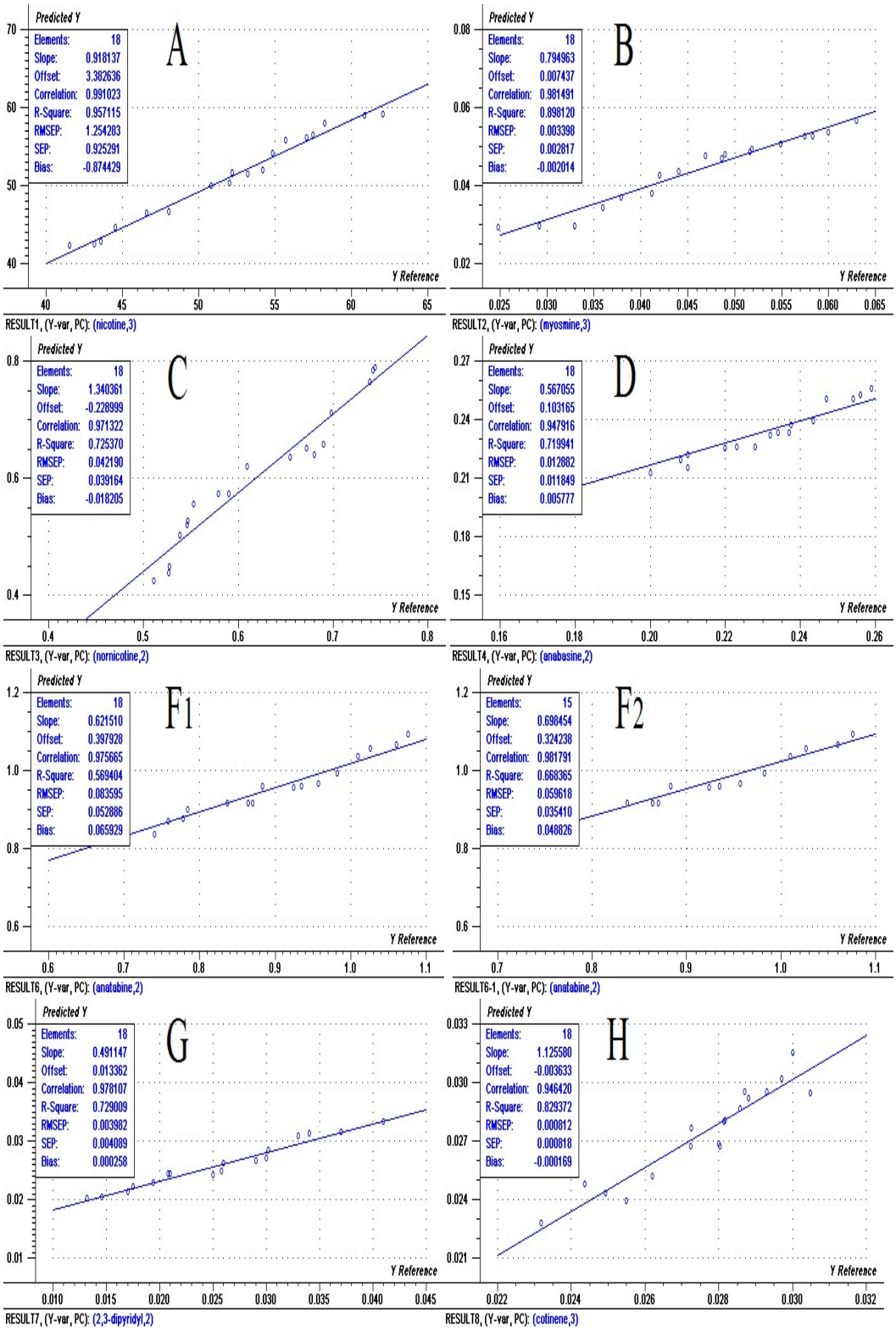



Fig.5. Validation of established prediction models for nicotine (A), myosmine (B), nornicotine (C), anabasine (D), anatabine (F1 and F2 indicate the model was validated by

eighteen and fifteen samples, respectively), 2,3-dipyridyl (G) and cotinine (H) through the examination of the other independent samples set.

## Graphical Abstract