RESEARCH ARTICLE

Investigation of the substitutability of tobacco raw materials based on the requirements of cigarette brands

Rong Ma[1], Yuanshang Liu[1], Xiaofu Shi[2], Pengyuan Zhang[1], Xiao Li[2], Yuntao Fan[1], Fuyun Cui[1], Longke Li[2], Qiannan Zhao[2], Yuan He[2, *]

[1]Technology Center, China Tobacco He'bei Industrial Co., Ltd., Shijiazhuang, Hebei, China. [2]Key Laboratory of Biotechnology in Tobacco Industry, College of Tobacco Science and Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China.

Received: September 9, 2025; **accepted:** December 29, 2025.

Reasonable substitution of tobacco raw materials is crucial for maintaining the stability of product quality. However, traditional evaluation methods mainly focus on sensory evaluations of smoking, which are limited by insufficient consistency and strong subjectivity, thereby restricting the accurate assessment of tobacco leaf substitutability. This research introduced Gray relational analysis method for the calculation of correlation degrees among different tobacco raw materials and standard samples. A matching system was constructed for the identification of substitutable tobacco leaves for the "Lotus" brand. The accuracy of the proposed method was verified by analyzing aromatic compound features of tobacco raw materials. Using weighted relational degree scores and clustering of key aromatic compounds, 17 tobacco raw materials were classified into three categories including materials with the correlation degree large than 87.76, accounting for 29.41%, materials with correlation range of 80.22 to 86.17, accounting for 47.06%, and those with correlation range of 77.07 to 80.05, representing 23.53%. Orthogonal partial least squares discriminant analysis (OPLS-DA) model based on aromatic compounds effectively identified intra-group similarities and inter-group differences among different samples. Further clustering analysis of tobacco raw materials using the criteria of $P < 0.05$ and VIP > 1 determined the key aromatic compounds, achieving 100% concordance between the classification results obtained from the two methods. Comprehensive evaluation of tobacco leaf quality using Gray relational analysis combined with clustering of key aromatic compounds could be applied to accurately obtain correlation degree interval for mutually substitutable tobacco leaves within the same category, thereby saving time in formulation design.

Keywords: Gray relational degree; orthogonal partial least squares discriminant analysis; tobacco raw materials; substitutability.

[*]**Corresponding author:** Yuan He, Key Laboratory of Biotechnology in Tobacco Industry, College of Tobacco Science and Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan 450001, China. Email: 2014095@zzuli.edu.cn.

Introduction

The core style of cigarette products is directly determined by the formulation of leaf group [1]. A leaf group formulation module consists of several tobacco raw materials with similar style characteristics combined within the formulation leaf group of the cigarette [2]. However, various factors such as ecological conditions and processing methods can influence these tobacco raw materials, potentially reducing the yield and quality of the product [3, 4]. Therefore, when a specific tobacco raw material is in short supply, substitute tobacco leaves should be promptly identified or alternative formulations with identical or similar style characteristics should be developed to maintain the stability of cigarette product quality [5, 6].

The most important part of finding appropriate tobacco raw materials and formulations mainly relies on empirical judgment and sensory evaluation or focuses primarily on cigarette formulation assessment [7, 8]. Currently, no predictive methods are available for raw materials based on the specific requirements of cigarette brands, and traditional evaluation systems have difficulty in meeting practical needs for objective and comprehensive assessment of the quality characteristics of tobacco leaves in terms of different dimensions. Grey relational analysis provides advantages such as minimal computational requirements, no restrictions on sample size or data distribution patterns, and quantitative results consistent with quantitative analyses [9]. Shen *et al*. performed Grey relational analyses to evaluate the influences of climatic factors during different growth periods on the main agronomic characteristics and yield of flue-cured tobacco [10]. Lin *et al*. conducted Grey relational analyses based on 9 trait indicators of sun-cured tobacco leaves for the selection of high-quality sun-cured tobacco cultivars [11]. Ran *et al*. utilized Grey relational method for the prediction of tobacco target spot disease [12]. However, whether used for evaluating the samples or building predictive models, when universally high relational degrees are obtained due to inherent natural commonalities among samples, Grey relational analyses cannot distinguish subtle differences among highly correlated samples. In addition, experience in the design and maintenance of cigarette products indicates that the intrinsic characteristics of tobacco raw materials should also be considered when looking for substitute tobacco leaves or formulations [13].

This research aimed to establish an objective and quantitative method for accurate prediction and screening of formulation modules capable of substituting for original tobacco raw materials to ensure the long-term stability of the core style and quality of cigarette products. To address the practical requirements of the cigarette brand "Lotus" for raw material substitutability, this research selected the aromatic component aspect of tobacco raw materials for the discrimination and classification of samples, which could help in determination of the appropriate range of relational degree for substitutable tobacco leaves, aiming to identify differences among tobacco leaves collected from different growing regions and decided whether they could be used to replace each other in actual production, thereby providing a theoretical basis to ensure cigarette brand quality stability.

Materials and methods

Tobacco leaf raw materials collection

A total of 35 tobacco leaf raw materials were randomly collected in 2023 by an expert team consisting of 7 tobacco leaf grading technicians from Hebei China Tobacco Industry Co., Ltd. (Shijiazhuang, Hebei, China), which consisted of two grades of middle leaf (C2F) and upper leaf (B2F) from Longyan, Fujian, China labelled as X_1 to X_3, Y_4, Y_{12}, Y_{17}; Baoshan, Yunnan, China labelled as X_4 to X_7, Y_3, Y_5, Y_6, Y_9; Chuxiong, Yunnan, China labelled as X_8, X_9, Y_7, Y_8; Qujing, Yunnan, China labelled as X_{10} to X_{12}, Y_{10}, Y_{11}, Y_{16}, Y_{18}; Qiannan Prefecture, Guizhou, China labelled as X_{13} to X_{16}, Y_1, Y_2, Y_{14}, Y_{15}; and Bijie, Guizhou, China labelled as X_{17}, Y_{13}.

Determination of physical and chemical indexes

Complete leaves from all tobacco leave grades were randomly selected and placed in a box under the temperature of 22 ± 1℃ and relative humidity of 60 ± 3% for 7 days to balance the moisture of tobacco samples. The thicknesses of 10 tobacco leaves from each producing area were measured by using thin slice thickness gauge and the average values were recorded. Manual stem extraction was performed on the intact leaves and only fine stems less than 0.5 mm in length were left at the tip of each tobacco leaf. The stems and tobacco leaves were then

Table 1. Sensory quality evaluation index and weight.

Indicators	Index	Weights	Scores				
			0 - 1	**1 - 2**	**2 - 3**	**3 - 4**	**4 - 5**
Aroma	Quality	0.20	Poor	Below average	Acceptable	Good	Excellent
	Volume	0.20	Less	Slight	Moderate	Fairly full	Full
	Off-flavor	0.10	Strong	Noticeable	Present	Slight	None
Smoke gas	Strength	0.10	Weak	Mild	Moderate	Strong	Intense
	Concentration	0.10	Bland	Light	Medium	Rich	Full-bodied
	Smoothness	0.15	Harsh	Rough	Slightly smooth	Fairly smooth	Smooth
Taste	Irritation	0.05	Strong	Noticeable	Present	Slight	None
	Aftertaste	0.05	Bitter	Lingering	Slightly lingering	Clean	Refreshing
	Sweetness	0.05	Faint	Slight	Noticeable	Pleasant	Distinct

respectively weighted by an electronic balance and stem percentage was calculated as stem weight/(stem weight + tobacco leaf weight) × 100%. Six discs with a radius of 0.75 cm were removed from the tips, leaves, and leaf bases of each intact leaf with a puncher and the weight of leaf mass was calculated as $\rho = m/(n \times s)$, where ρ was leaf mass weight (g/m^2). *m* was disc weight (g). *n* was the number of discs/pieces. *s* was disc area (m^2). The balanced intact leaves were weighted and placed into DHG-9145A electric air blower drying oven (Shanghai Yiheng Technology Co., Ltd., Shanghai, China) for 2 hours after the temperature reached 100 ± 2℃. The dried tobacco leaves were placed in a dryer, cooled to room temperature, then weighted and recorded as dry weight. Equilibrium moisture content was calculated as (fresh weight-dry weight)/dry weight × 100%. The main veins of intact tobacco leaves were extracted. Strips with width of 1.5 cm and length of 15 cm were cut, and tensile strength was measured by using ZKW-3 tobacco sheet tensile tester (Sichuan Changjiang Paper Instrument Co., Ltd., Luzhou, Sichuan, China). The conventional chemicals in tobacco leaves were detected according to China tobacco industry standards YC/T 159-2019 [14], YC/T 160-2002 [15], YC/T 161-2002 [16], YC/T 217-2007 [17], and YC/T 162-2011 [18].

Sensory evaluation

Sensory evaluation of fermented flue-cured tobacco leaves was performed according to the Chinese tobacco industry standard YC/T 530-2015 and previous published method by Xu *et al.* [19, 20] to evaluate aroma characteristics including volume, quality, and off-flavour; smoke features including concentration, strength, and smoothness; and mouthfeel including aftertaste, irritation, and sweetness. Seven expert evaluators from the Technical Centre of Hebei China Tobacco Industrial Co., Ltd (Shijiazhuang, Hebei, China) independently evaluated and scored the samples. The scoring ranges and weights of each evaluation attribute obtained during sensory-based formulation work were shown in Table 1 with the mean score calculated after assessment. The overall sensory quality score was calculated as follows:

Total sensory score = (aroma quality × 0.20 + aroma volume × 0.20 + off-flavour × 0.10 + strength × 0.10 + concentration × 0.10 + smoothness × 0.15 + irritation × 0.05 + aftertaste × 0.05 + sweetness × 0.05) × 100 / 5.

Determination of aromatic compound

After conditioning tobacco powder samples for 48 h in a humidity chamber under constant temperature of 22 ± 1°C and relative humidity of 60 ± 3%, 25 g aliquots of sample were weighted and placed into a 1,000 mL flat-bottom flask. Then, 25 g sodium chloride and 350 mL deionized water were added to the samples. The obtained mixtures were heated using an electric heating mantle to maintain gentle boiling. At the other end of the apparatus, a 100 mL flask containing 50 mL dichloromethane was placed in a 60°C water bath. A 1°C low-temperature cooling circulation pump was applied as condenser water source for simultaneous distillation-extraction apparatus. The distillation-extraction process performed for 2.5 h. The obtained extract was dehydrated overnight using anhydrous sodium sulphate with 1 mL phenethyl acetate being added as internal standard. The obtained mixture was concentrated in a 60°C water bath to 1 mL before being transferred to a chromatography vial and subjected to gas chromatography-mass spectrometer (GC-MS) analysis using Agilent 7890B/5977 GC-MS (Agilent, Santa Clara, California, USA) [21]. Separation was performed using an HP-INNOWax capillary column (Agilent, Santa Clara, California, USA) with dimensions of 30 m × 250 μm × 0.25 μm and split ratio of 5:1 under the temperature program of initial temperature 50°C for 2 min, heated to 280°C at 4 °C/min for 20 min. The temperature of injection port was 280°C. Mass spectrometry analysis was carried out in electron ionization (EI) mode with scan range of 30 - 700 m/z, ion source temperature of 230°C, and quadrupole temperature of 150°C.

Analysis of key aromatic components

SIMCA 14.1 software (Umetrics, Umea, Sweden) was utilized to perform orthogonal partial least squares discriminant analysis (OPLS-DA). The key aromatic components were identified based on the combined criteria of univariate t-test results ($P < 0.05$) and variable importance in projection (VIP > 1) in OPLS-DA model, which was consistent with established screening standards [22].

Computational for comprehensive weights

Criteria importance though intercriteria correlation (CRITIC) weighting method is an objective weighting approach, which comprehensively investigates indicators based on their conflicting character and contrast intensity. The objective weights of each indicator in this study were calculated by using CRITIC as follows.

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij}-\bar{x}_j\right)^2}{n-1}} \quad (1)$$

$$C_j = \sigma_j \sum_{i=1}^{n}\left(1 - r_{ij}\right) \quad (2)$$

$$W_j = \frac{C_j}{\sum_{j=1}^{n} C_j} * 100\% \quad (3)$$

where σ_j was standard deviation. r_{ij} was correlation coefficient between indicators *i* and *j*. C_j was the information content contained in indicator *j* such that a larger value of C_j indicated greater information content in indicator *j* signifying its importance within the system. W_j was the weighting coefficient of indicator *j*. However, assigning weights solely based on objective data might overlook the multidimensional nature and complexity of the problem. Therefore, it was essential to incorporate the practical significances of indicators alongside subjective expert judgment, specifically by applying analytic hierarchy process (AHP) to obtain subjective weights. This study constructed a judgment matrix based on 1 - 9 scaling method by comparing the relative importances of the indicators (Table 2). After obtaining judgment matrix, the subjective weight of each indicator was calculated as follows.

$$M_i = \prod_{j=1}^{n} b_{ij} \quad (4)$$

$$\bar{W}_i = \sqrt[n]{M_i} \quad (5)$$

$$W_i = \frac{\bar{W}_i}{\sum_{j=1}^{n} \bar{W}_i} * 100\% \quad (6)$$

where M_i was the product of the elements in each row of judgment matrix. b_{ij} was the element in the *i*-th row and *j*-th column. $\bar{W}_i$ was the *n*-th root value of the row product. W_i was eigenvector. The obtained judgment matrix was subjected to randomness and consistency tests using the equations $CI = \frac{\lambda_{max}-n}{n-1}$ and $CR = \frac{CI}{RI}$ where λ_{max} was the maximum eigenvalue. The consistency ratio (CR) of the judgment matrix was acceptable when its value was lower than 0.10. Comprehensive weights were calculated by integrating the objective weights derived from CRITIC method with subjective weights obtained from AHP (Table 3) using multiplicative normalization method as follows.

Table 2. The judgment matrix signs of AHP and their implications.

Scale	Indications
1	factor a and factor b were of equal importance
3	factor a was moderately more important than factor b
5	factor a was strongly more important than factor b
7	factor a was very strongly more important than factor b
9	factor a was extremely more important than factor b
2,4,6,8	intermediate values between two adjacent judgments above
Reciprocal	If the comparison ratio for the importance of factors a to b was M_{ab}, the ratio for b to a was M_{ba} = 1/M_{ab}

Table 3. Values of RI in the AHP.

n	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

$$W_i^3 = \frac{W_i^1 W_i^2}{\sum_{i=1}^{n} W_i^1 W_i^2} \quad (7)$$

where W_i^3 was comprehensive weight. W_i^1 was the weight obtained from CRITIC method. W_i^2 was the weight determined by AHP.

Grey relational analysis

Assuming that Grey relational analysis was suitable for small sample sizes and did not impose strict requirements on data distribution patterns, this research adopted this method for the calculation of relational degrees among raw materials from different tobacco-growing regions and the standard sample, aiming to evaluate their substitution potentials with the calculation as follows.

$$Z_{ij} = \frac{Y_{ij}}{\bar{Y}_j} \quad (8)$$

$$\delta_{ij} = \frac{1}{n}\sum_{j=1}^{n} \frac{min_i min_j |Z_{0j} - Z_{ij}| + \rho max_i max_j |Z_{0j} - Z_{ij}|}{|Z_{0j} - Z_{ij}| + \rho max_i max_j |Z_{0j} - Z_{ij}|} \quad (9)$$

where Z_{ij} was the dimensionless value of the *j*-th indicator in the *i*-th tobacco-growing region after processing. Y_{ij} was the measured value of the *j*-th indicator in the *i*-th tobacco-growing region. $\bar{Y}_j$ was the mean value of a given evaluation indicator. δ_{ij} was the relational degree between the raw material from the *i*-th tobacco-growing region and standard sample X_0 with respect to the *j*-th evaluation indicator. $\rho \in (0, 1)$ was the distinguishing coefficient, typically set at $\rho = 0.5$.

Results

Overall distribution characteristics of quality assessment indicators of tobacco leaf raw materials

The distribution of indicators for the evaluation of the quality of tobacco leaf raw material demonstrated that variation coefficients for reducing sugar, equilibrium moisture content, and nitrogen-alkaloid ratio were relatively small with the values of 6.25%, 9.16%, and 9.77%, respectively, which indicated that most tobacco raw materials were very similar in terms of these three indicators with relatively concentrated data distributions and small fluctuation ranges [23]. In contrast, the variation coefficients of sugar-alkaloid ratio, total nitrogen, and tensile strength were larger with the values of 48.49%, 35.56%, and 33.19%. respectively, indicating substantial differences among tobacco leaf samples in these three indicators (Table 4).

Prediction of alternative raw materials based on tobacco leaf quality evaluation indicators

(1) Developing ideal evaluation indices in Grey relational analysis

Grey relational coefficients among the quality evaluation indicators of the tobacco raw

Table 4. Distribution of evaluation index content of tobacco raw material quality.

Index	Mean ± SD	Kurtosis	Skewness	Minimum	Maximum	CV (%)
Equilibrium moisture content (%)	19.97 ± 1.83	3.93	1.72	17.31	21.94	9.16
Stem content rate (%)	26.00 ± 2.78	-1.26	0.22	22.65	31.20	10.69
Leaf quality is heavy (g/m^2)	8.73 ± 1.03	0.57	1.05	7.50	11.49	11.80
Blade thickness (μm)	100.63 ± 15.69	0.41	0.53	86.28	123.82	15.59
Pull (N)	2.60 ± 0.86	0.41	0.29	1.83	3.87	33.08
Nicotine (%)	2.58 ± 0.72	0.50	0.29	2.07	3.06	27.91
Total nitrogen (%)	2.24 ± 0.76	-1.36	0.15	1.76	2.83	33.93
Reducing sugar (%)	17.60 ± 1.10	0.54	0.26	15.11	20.08	6.25
Potassium (%)	2.92 ± 0.74	1.06	1.09	1.92	4.99	25.34
sugar-alkali ratio	7.10 ± 2.53	2.15	1.42	3.97	15.02	35.63
Nitrogen-to-alkali ratio	0.94 ± 0.09	-1.00	0.59	0.79	1.11	9.77
Potassium to chloride ratio	6.86 ± 1.73	1.54	0.05	3.99	9.98	25.22
Aroma quality	3.28 ± 0.59	0.22	0.23	1.80	4.20	17.99
Amount of aroma	3.13 ± 0.49	0.90	0.39	1.80	4.20	15.65
Miscellaneous gas	2.48 ± 0.62	0.18	0.83	1.80	4.20	25.00
Irritating	2.77 ± 0.46	0.33	1.06	1.80	3.20	16.61
Aftertaste	3.17 ± 0.54	0.22	0.21	1.80	4.20	17.03

materials collected from various producing regions and the ideal reference values showed that, in the predictive evaluation system developed based on Grey relational analysis, each indicator required the construction of a reference sequence for comparative analyses (Figure 1). According to Ma *et al*. [24], when determining the ideal reference values for various quality indicators, the maximum measured values for potassium content, sensory quality indicators, and potassium-chloride ratio (higher values were more desirable) were adopted as ideal reference values. Conversely, stalk rate was viewed as an indicator where smaller values indicated higher quality and represented by the minimum measured value. For the remaining indicators with both upper and lower limits (moderate indicators), the ideal reference values were determined based on quality requirements for tobacco raw materials from Hebei China Tobacco (Shijiazhuang, Hebei, China). Considering the raw material indicator values from various tobacco-growing regions as child sequences and the ideal values of standard samples as the parent sequence, the subsidiary sequence with the largest Grey relational coefficient had the closest relationship with the parent sequence.

(2) Calculation of the weights of quality evaluation indicators

The comprehensive weights obtained from AHP combined with the coefficient of variation method showed that, among the weights assigned to chemical composition, sensory quality, and physical properties, the highest-ranking indicators were the quality of aroma, sugar-alkali ratio, and pull with weights of 11.31%, 10.07%, and 5.89%, respectively. On the other hand, the indicators with the lowest weights were potassium content, irritancy, and leaf thickness with the values of 2.18%, 6.39%, and 1.64%, respectively (Figure 2). Overall, the weights of aroma volume, aroma quality, and off-flavor constituted a larger proportion of the total value, while potassium, equilibrium moisture content, and leaf thickness had comparatively lower weights. The overall trend of weight distribution was in line with the findings of Wang *et al*. [25].

(3) Calculation of relational degrees

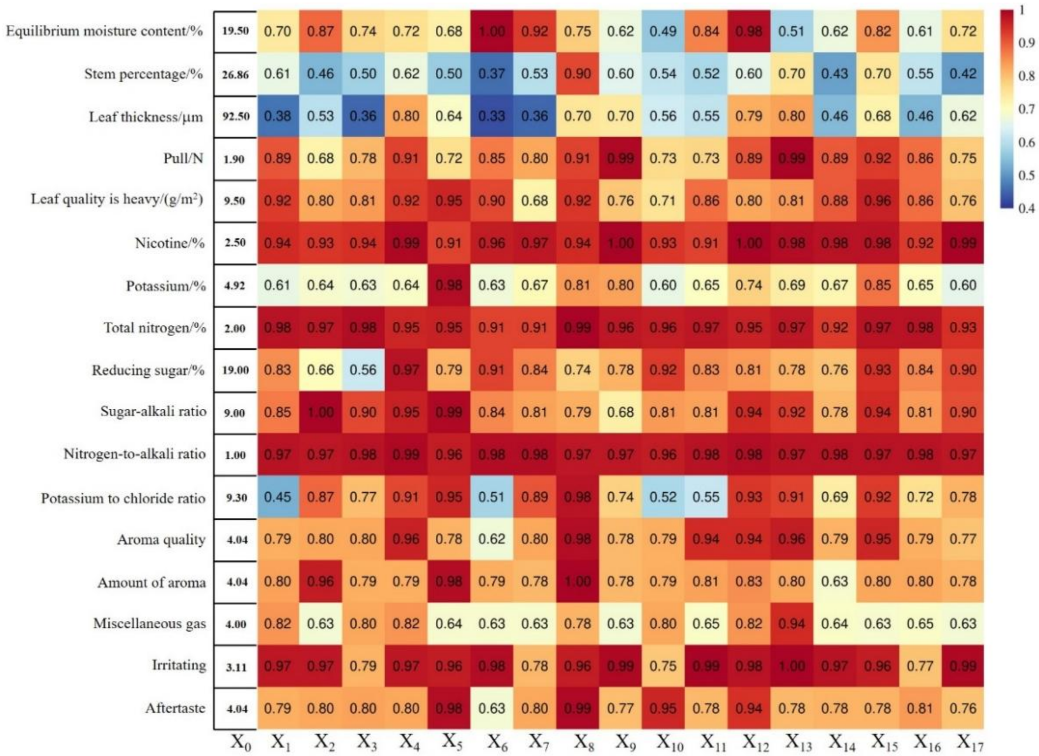

Figure 1. Correlation number and reference value of standard products for raw material quality evaluation index of tobacco leaf.

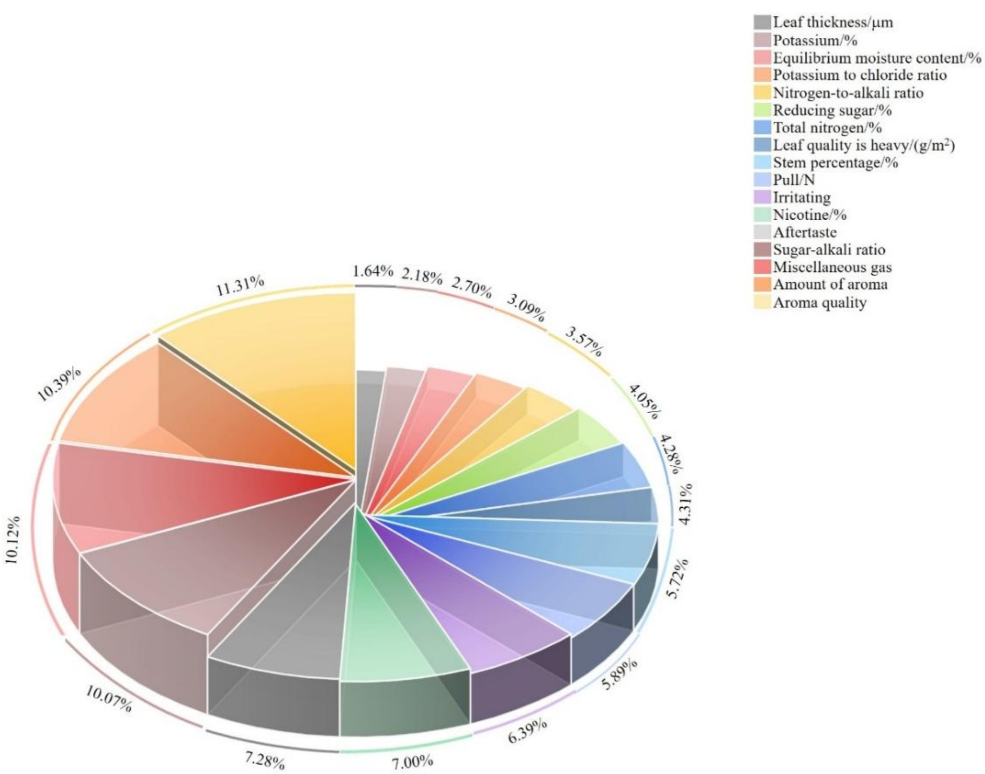

Figure 2. Comprehensive weight of quality evaluation indicators.

Weighted relational degrees among all tobacco leaf raw materials and the standard sample were significant with values of greater than 75.00. Specifically, X8 exhibited the highest weighted relational degree of 91.68, being the closest to the standard sample. The weighted relational degrees of the remaining samples were in the range of 77.07 to 89.57 with the order as X12 (89.57) > X13 (89.04) > X4 (88.87) > X15 (87.76) > X5 (86.17) > X1 (82.76) > X2 (82.75) > X11 (81.41) > X3 (81.25) > X10 (80.93) > X16 (80.40) > X9 (80.22) > X7 (80.05) > X17 (80.00) > X14 (77.77) > X6 (77.07). Although all samples obtained certain correlation levels, relatively small differences were witnessed in relational degrees between adjacent samples, reflecting high degrees of common characteristics among the samples. Therefore, when selecting alternative formulations or tobacco leaves, it was important to consider not only the quality standards of the tobacco raw materials but also the intrinsic characteristic of the tobacco leaves to help in accurately classifying the samples.

(4) Difference analysis of aromatic components in tobacco raw materials

Aromatic components form the material basis of the characteristic flavor profile of tobacco. Differences in their concentrations and varieties lead to variations in the distinctive flavors of different samples. However, not all aromatic components directly determine tobacco leaf characteristic flavor, as most may not exert practical influence due to minimal olfactory contribution or low concentrations. Comprehensive analysis of all aromatic components could introduce bias. Therefore, the key aromatic components affecting tobacco classification in this study were screened using thresholds of $P < 0.05$ and VIP > 1.00. The results showed that 14 intergroup-differentiating components were identified, which critically determined cigarette aromatic profiles [26]. Most components had positive contribution to sensory quality by enriching fragrance, decreasing smoke irritation [27], and enhancing aromatic harmony [28]. In OPLS-DA model, dependent variable fit index (R_y^2) was 0.91, independent variable fit index (R_x^2) was 0.87, and model predictive index (Q^2) was 0.80, all of which exceeded 0.50, indicating acceptable model fitting (Figure 3A). After 200 permutation tests and 7-fold cross-validation, the intercept of Q^2 regression line on y-axis was -0.69 (< 0), confirming the validity of the proposed model without overfitting (Figure 3B). Under valid model conditions, tobacco samples within the same category were clustered closely, demonstrating strong internal similarity, while distinct separation between categories indicated unique characteristics of aroma composition. Among them, X2 and X8 significantly deviated from the group centroid (Figure 3C). Research has indicated that this deviation could be attributed to intra-group heterogeneity due to aromatic components that play key role in distinguishing tobacco leaf raw materials. To further explore the effect mechanism of heterogeneity on substitutability within sample groups, given that higher VIP values indicated greater discriminatory power and sample influence, hierarchical cluster analyses were conducted using 14 key aromatic components for the optimization of classification accuracy. Samples were classified into three distinct clusters including cluster 1 that comprised 5 production regions of X8, X12, X4, X15, and X13, accounting for 29.41% of the total value; cluster 2 that aggregated 8 production regions of X5, X10, X11, X2, X1, X16, X9, and X3, representing 47.06%; and cluster 3 that contained 4 production regions of X7, X6, X14, and X17, accounting for 23.53%, exhibiting the closest similarity to cluster 2 in overall quality characteristics of aroma (Figure 3D).

Results of practical substitution based on sensory quality of tobacco raw materials

The sensory evaluation qualities of tobacco leaf raw materials were compared. Among them, X13 was free of off-odors, while X6 presented low aromatic quality and lingering aftertaste, and X14 had a slightly deficient volume of aroma. In addition, based on comprehensive quality assessment of X8, its well-balanced chemical composition played a key role in its high overall

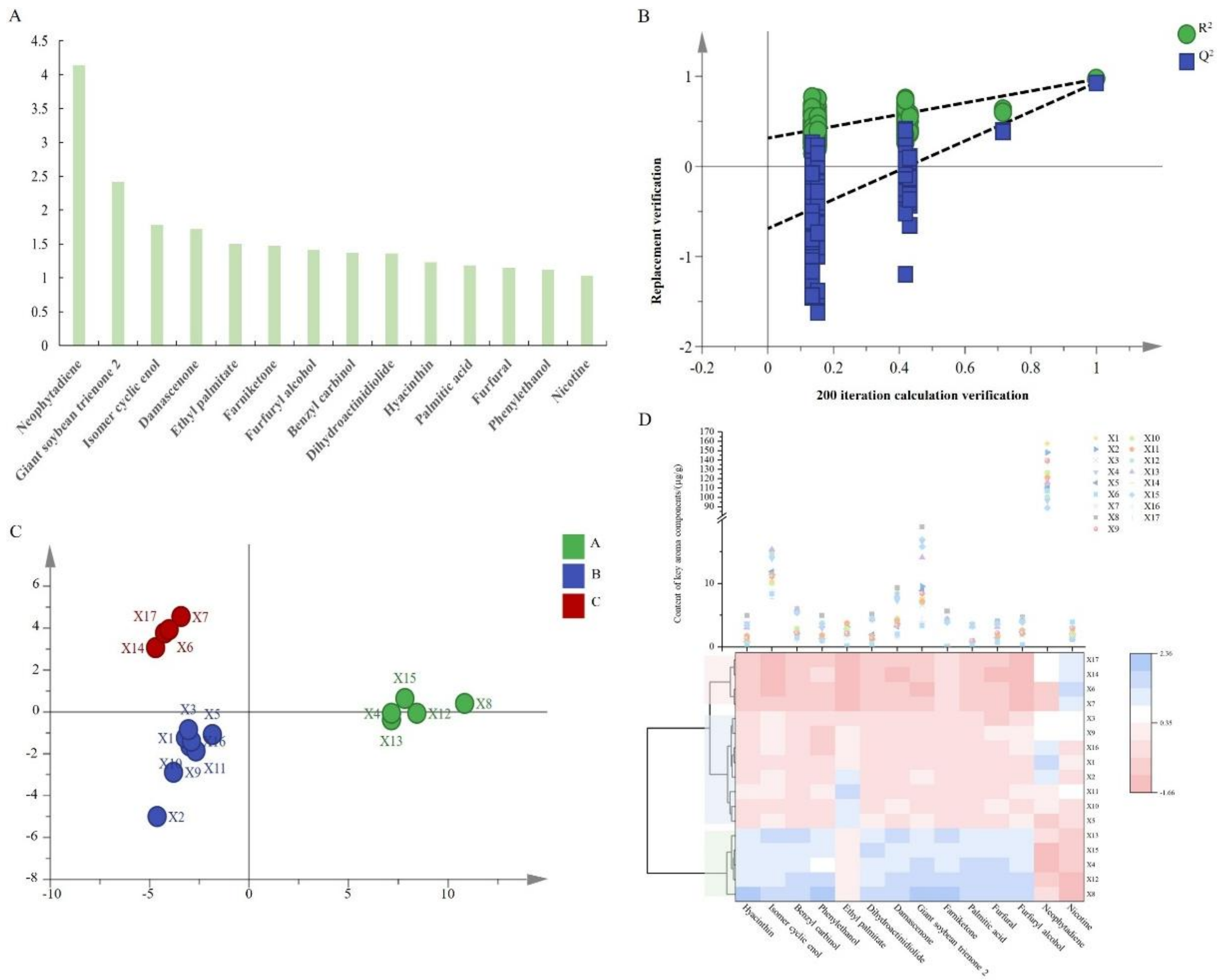

Figure 3. Classification model based on aroma components of tobacco raw materials. **A.** VIP value diagram of key aroma components. **B.** Cross-validation results of OPLS-DA model. **C.** OPLS-DA score chart of aroma components in tobacco raw materials. **D.** Content of key aroma components and clustering results.

sensory evaluation score of 72.80. However, the total sensory evaluation score of X6 was 51.93, which was remarkably lower than those of tobacco leaves from other regions. The prediction results of the grey relational analysis were verified using sensory quality scores. The obtained results revealed that the actual sensory evaluation rankings of X3, X9, X10, and X16 were 13th, 12th, 10th, and 11th, respectively, while the priority order determined by Grey relational analysis was 10th, 11th, 12th, and 13th, indicating some discrepancies. Among all 17 tobacco leaf raw materials, the overall consistency rate for substitution priority ranking was 76.47%.

Validation of the universality of the substitution system

To validate the universal applicability of the proposed substitution system, tobacco leaf raw materials from 18 known growing regions with grade B2F were evaluated using the same experimental procedure, for which substitution category and priority were predetermined. Weighted relational degrees for B2F tobacco leaves ranged from 80.08 to 88.79. Consistency rate of 100% was obtained between the predicted and actual substitution categories, indicating that the predictive system exhibited excellent classification performance and could accurately distinguish similarities among different tobacco leaf raw materials.

Discussion

Construction of Grey relational model

Sustainable and healthy development of cigarette brands mainly relies on stable supply of high-quality flue-cured tobacco raw materials. Continual improvement and optimization of leaf blending techniques are essential for decreasing the dependence of cigarette manufacturers on scarce and characteristic tobacco types [29]. Extensive research has been conducted using various analytical methods to ensure the stability of cigarette quality, among which Grey relational analysis demonstrated favorable applicability in the evaluation of the quality of tobacco leaves [30]. Compared to traditional principal component analysis, Gray relational analysis is more suitable for dynamic system analyses and offers superior predictive performance [31]. Guided by the practical requirements of the cigarette brand "Lotus", this study advanced the evaluation of tobacco leaf substitutability by shifting from a reliance on sensory quality scores for assessing similarity to the application of Grey relational analysis. This innovative approach allowed for a more systematic and scientific determination of substitution priorities among tobacco leaves. The results indicated that X8 was the most appropriate high-quality tobacco leaf. Correlation degrees among other tobacco samples and the standard sample generally exceeded 75.00. High correlation degrees might be attributed to common quality features among various tobacco leaves, which inherently resulted in some degree of natural association among the samples [32]. Hence, selection of substitute leaves or formulations should not only rely on quality standards but also take into account the distinctive characteristics and styles of tobacco to perform a comprehensive selection of substitution targets.

Accuracy verification of Grey relational system

Shaped through complex interactions, aromatic compounds highlight the characteristic flavors of tobacco and play a key role in tobacco quality [26, 33]. This study identified 14 key aromatic components by using OPLS-DA model with significant inter-group differences and the standard of VIP > 1.00. Cluster analysis of these 14 influential aromatic components, combined with the intrinsic features of tobacco raw materials, classified the samples into three categories. Within OPLS-DA model, however, samples X2 and X8 deviated from the centers of their respective groups. Further analyses showed that this deviation could be attributed to remarkable differences in palmitic ester and neophytadiene contents in second-category raw materials. It should be noted that only X2 presented higher concentrations of these two aromatic compounds compared to other samples in the group, which likely caused its deviation. Although the trend of the data obtained for the 14 key aromatic components was relatively consistent within the first category, X8 and other tobacco samples in its group presented overall differences in the contents of aromatic compounds. These clustering results highlighted heterogeneity within OPLS-DA classification and provided a reasonable explanation for this variability [34].

Universal verification of Grey relational system

Compared to methods solely relying on sensory evaluation for the substitution of tobacco formula, the substitution system proposed in this research was more comprehensive and accurate. Its high priority consistency rate indicated robust applicability. In addition, utilization of different tobacco grades for the validation of the universal applicability of the system confirmed that substitution categories based on weighted relational interval division aligned with actual classification results. Therefore, this classification framework was employed for the determination of the weighted relational degrees of potential substitute tobacco, facilitating rapid allocation within the same category during tobacco substitution processes. This technique offered a convenient and rational basis for formulation design. While Grey relational analysis demonstrated higher accuracy in small-sample research, future research should adopt various analytical techniques and larger sample size to further improve and refine this substitution system.

Acknowledgements

This research was supported by the Science and Technology Research Project of Henan Province (Grant No. 232102230107) and the Science and Technology Project of China Tobacco Hebei Industrial Co., Ltd (Grant No. 2023130000340255).

References

1. Gao X. 2022. Present situation and technical discussion of Chinese cigarette processing technology. New Industrialization. 12(9):76-79.
2. Tian Z, Xu Q, Miao C, Sun J, Li S, He H, *et al*. 2023. Exploration of dynamic balance prediction and distribution methods of tobacco leaf raw materials based on formulation functions. Chin J Tob Sci. 29(1):1-10.
3. Li L, Xu Y, Liu B, Wang C, Bi Y, Liu X, *et al*. 2015. Analysis of aroma compounds in flue-cured tobacco leaves from different ecological regions. Tob Sci. 36(3):1-7.
4. Tan G, Liu Y, Peng S, Yin H, Meng D, Tao J, *et al*. 2021. Soil potentials to resist continuous cropping obstacle: Three field cases. Environ Res. 200(suppl-C):111319.
5. Li Q, Chen K, Deng X, Guo S, Chen H, Zhong H, *et al*. 2018. Substitution method of tobacco leaves based on analysis of difference degree of tobacco pyrolysis. Tob Technol. 51(8):77-84.
6. Hu J, Yang Y, Liu M, Liu X, Wang B, Niu Y, *et al*. 2020. Chemical composition differences in cut-leaf re-dried cigarettes of different sizes. Journal of Northwestern Agriculture and Forestry University. 48(9):89-96.
7. Jiang M, Zheng W, Huang F, Zhang X, Yang D, Yang Y, *et al*. 2013. Evaluation of the intrinsic quality conformity of Dali K326 tobacco leaves based on brand-oriented assessment. Hubei Agricultural Sciences. 52(13):3068-3071, 3074.
8. Feng T, Ma L, Ding X, Yang N, Xiao X. 2007. Intelligent techniques for cigarette formula design. Math Comput Simul. 77(5):476-486.
9. Jiang W, Wang Z, Chen B, Ma J, Bao N, Chen G, *et al*. 2025. Interactive effects of planting patterns combined with integrated nutrient management on maize production, water-nitrogen productivity and soil organic carbon fractions. BMC Plant Biology. 25(1):545-545.
10. Shen H, Yang K, Tang Z, Zeng W, Huang F, Wan R, *et al*. 2024. Effect of transplanting period on upper leaf quality and economic traits of flue-cured tobacco in Longyan. Tob Sci. 45(6):55-64.
11. Lin S, Wang H, Zhao W, Shen Y, Zong H, Liu S, *et al*. 2019. Screening of high-quality sun-cured tobacco cultivars based on clustering of nicotine, total sugar, and agronomic traits combined with grey relational analysis. Tob Sci. 52(4):17-24.
12. Ran Y, Jin Y, Wang Z, Cheng X, Sun J, Luo J, *et al*. 2024. Construction and digital application development of a prediction model for tobacco target spot disease. Plant Medicine. 3(4):40-49.
13. Nong S, Luo G, Yuan Q, Huang J, Huang B, Lu Y, *et al*. 2025. Study on characters of different cigar varieties in baise. Acta Tobacco Sinica. 48(3):397-404.
14. Tobacco and tobacco products-Determination of water-soluble sugars-Continuous flow method: YC/T 159-2019. China Standard Press. 2019.
15. Tobacco and tobacco products-Determination of total phytoalkaloids-Continuous flow method: YC/T 160-2002. China Standard Press. 2002.
16. Tobacco and tobacco products-Determination of total nitrogen-Continuous flow method: YC/T 161-2002. China Standard Press. 2002.
17. Tobacco and tobacco products-Determination of potassium-Continuous flow method: YC/T 217-2007. China Standard Press. 2007.
18. Tobacco and tobacco products-Determination of chlorine-Continuous flow method: YC/T 162-2011. China Standard Press. 2011.
19. Sensory evaluation method of quality, style and characteristics of flue-cured tobacco leaves: YC/T 530-2015. China Standard Press. 2015.
20. Xu X, Yang Z, Li Z, Kong D, Yang Q, Yang X, *et al*. 2024. Quality status of tobacco leaves in Kunming production area and screening of new reserve flue-cured tobacco varieties. Journal of Kunming University. 46(3):1-10.
21. Li Y, Wang Y, Yuan D, Li Y, Zhang L. 2020. Comparison of SDE and SPME for the analysis of volatile compounds in butters. Food Sci Biotechnol. 29(1):55-62.
22. Zhu C, Zhou Y, Xie Q, Pan Y, Zhao Y, Liu H, *et al*. 2025. Integration of E-tongue, GC-IMS and chemometrics for unraveling the disparities in volatile profiles of optimized jujube craft beers. J Chromatogr. 1760(000):466264.
23. Lu Y, Mao X, Shen R, Du P, Qi H, Bai X. 2022. Analysis and evaluation of nutritional quality of introduced potato varieties in northern Shanxi Province. Chinese Potato Journal. 36(5):385-395.
24. Ma B, Zhang Y, Sun Y, Huang P, Chen X, Sun X, *et al*. 2023. Study on the utilization and positioning of raw materials in initial tobacco formulation modules. Journal of Yunnan Agricultural University. 38(2):324-332.
25. Wang Y, Xie J, Li Z. Tobacco planting regions in China: A mapping. Beijing: Science Press. 2010.
26. Yu Y, Yang Y, Jia J, Zhou H, Qiu Y, Sun M, *et al*. 2025. Changes in microbial composition during flue-cured tobacco aging and

their effects on chemical composition: A review. Bioresour Bioprocess. 12(1):43.

27. Hu W, Cai W, Zheng Z, Liu Y, Luo C, Xue F, *et al*. 2022. Study on the chemical compositions and microbial communities of cigar tobacco leaves fermented with exogenous additive. Sci Rep. 12(1):19182-19182.
28. Zou L, Su J, Xu T, Ji X, Wang T, Chen Y, *et al*. 2023. Untargeted metabolomics revealing changes in aroma substances in flue-cured tobacco. Open Chem. 21(1):3833.
29. Zhou J, Zhang J. 2008. Constructing a guaranteed system of high-quality characteristic tobacco raw materials for Chinese-style cigarettes as a strategic choice under new circumstances. Tob Sci. 14(1):42-46, 57.
30. Wang Z, Liu K, Yao J, Tian Y, Yang J, Zhao W, *et al*. 2017. Comprehensive evaluation of Zimbabwe flue-cured tobacco varieties based on gray relational analysis. Tob Sci. 38(4):92-96.
31. Jin X, Pan T, Yu H, Wang Z, Cao W. 2025. Electricity load forecasting method based on the GRA-FEDformer algorithm. Energies. 18(15):4057.
32. Rong L, Xue Z, Deng C, Luo K, Gao C, Li S, *et al*. 2024. Ecological adaptability and comprehensive evaluation of seven tobacco varieties planted in Lijiang, Yunnan. Chinese Agricultural Bulletin. 40(22):17-23.
33. Hou B, Liu Z, Liu Y. 2020. Effects of altitudes on chemical components and aroma components in flue-cured tobacco leaves with different genotypes. Journal of Jiangxi Agriculture. 32(10):87-92.
34. Luo T, Qu W. 2024. Clustering algorithm of local direction centrality measure for classification of agricultural machinery track field road. Comput Eng Appl. 60(23):303-313.