

Origin Traceability of Tobacco Leaves by Multi-elements Analysis Using Inductively Coupled Plasma Atomic Emission Spectroscopy and Cluster Analysis

Hua Wang

Abstract—An Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) method was developed for determination of 16 elements (As, Se, Zn, Pb, Cd, Ba, Fe, B, Mn, Cr, Mg, V, Ca, Cu, Al, K) in tobacco leaves. Collected 36 samples from Peru, Zimbabwe, and China were analyzed. The results showed that the content of Cd, Pb, As, Cr and other heavy metals in all samples did not exceed the limit of China national regulation. Principal component factor (PCA) analysis indicated that the studied 16 elements are not well correlated. The variations of elements in tobacco leaves was attributed to the difference of environmental soil rather than the result of tobacco secondary metabolism. 36 samples can be approximately divided into 10 groups. This mathematical model can serve as a traceability tool for tobacco leaf products. This method may facilitate the tobacco selection and quality identification.

Index Terms—tobacco leaf, element analysis, ICP, principal component analysis, origin traceability.

I. INTRODUCTION

Tobacco is recognized as an important hazard to human health [1]. On the other hand, its related industries are important economic components and tax sources. Tobacco is also a raw material for natural products [2]. Therefore, the selection of low harmful tobacco has become the concern of scientific researchers as well as reducing harmful substances in tobacco and improving the quality of related product. The qualitative and quantitative analysis of chemical components in tobacco leaves is the basis of such research.

In the late 1980s and early 1990s, scholars have reported harmful ingredients in tobacco leaves [3]. The list of components includes seven heavy metal elements and 107 harmful components in cigarette smoke. The content of inorganic elements in different parts of tobacco plants was always similar when tobacco plants are from same geographic region [4,5]. On the contrary, the metal content of tobacco varies significantly among different regions [6,7]. One main reason is that tobacco is closely related to soil element background and ecological environment conditions [8,9]. For example, the level of heavy metals in the soil of the planting area is the main factor affecting the mass fraction of heavy metals in tobacco [10].

Many techniques have been developed to study the

geographical origin of plant or animal product. For example, near Infra-Red spectroscopy was applied for geographical origin study of cocoa beans [11]. Stable isotope analysis was carried out to trace the origin of velvet antlers [12]. GC-MS was combined with multivariate analysis to determine the geographical origin of Chinese herbal medicine [13]. Combination of multi-techniques was also tried for traceability of milk [14]. Regardless of the above techniques, one of the most important methods of geographical traceability is based on element analysis[15,16]. Geographical origins were traced successfully based on elements and multivariate analysis for rice [17], soybean [18], potato [19,20], croaker [21], sea cucumbers [22] and so on.

The existing reports on regional research on heavy metals may not be representative due to the limited number of heavy metal objects studied[23] or the small sample size and the limited of tobacco geographic origin [5,23]. Other approaches were not based on element and needed complicated analytical techniques [14,24–26].

In this study, 16 elements (As, Se, Zn, Pb, Cd, Ba, Fe, B, Mn, Cr, Mg, V, Ca, Cu, K, Al) in tobacco leaves from 36 major tobacco growing areas around the world were qualitatively and quantitatively investigated by inductively coupled plasma atomic emission spectrometry (ICP-AES). The level and distribution of heavy metal content in different regions were analyzed. Then the principal component analysis tool is used to study the traceability and regional correlation of tobacco products. This clustering analysis method can also be used to distinguish the zoning of tobacco flavor characteristics.

II. EXPERIMENTAL

A. Reagents and standards

The 36 samples of tobacco leaves were collected from 8 regions. Three of them were from Zimbabwe (Africa), three from Peru (South America) and the rest from China. Among the samples from China, 15 were from Yunnan Province, 3 from Hunan, 3 from Guizhou, 3 from Henan, 3 from Anhui and 3 from Xinjiang. The standard stock solutions of As, Se, Zn, Pb, Cd, Ba, Fe, B, Mn, Cr, Mg, V, Ca, Cu, K and Al were all purchased from the National Nonferrous Metals and Electronic Materials Analysis and Testing Center. The concentration of each solution was 1000 mg/L. Nitric acid is guaranteed reagent, and other chemical reagents are analytical

reagent.

All solutions used were prepared with ultrapure water ($R > 18 \text{ M}\Omega/\text{cm}$; $\lambda < 0.001/\text{cm}@254 \text{ nm}$). All glassware used was soaked in 1% nitric acid for 24 hours to remove interference ions..

B. Instrument and operating parameters

The plasma emission spectrometer used was a Perkin Elmer (Waltham, MA, USA) Optima 8000DV. Instrument parameters: RF power 1350 W; 16 center detection wavelengths for 16 elements; argon pressure of injection atomizer was 28 psi; the injection flow rate was 1.5 mL/min. The auxiliary gas flow rate was 1.0 L/min and the carrier gas flow rate was 1.12 L/ min. The single peak detection was repeated.

The microwave digestion instrument was model in ETHOS A (Milestone, Italy). The milli-Q Synthesis (Millipore, France) was used to provide ultrapure water. The electronic balance (precision 0.1 mg) was model in BSA224S (Cedorius Scientific Instruments Co., Ltd., China).

C. Sample preparation

The tobacco leaves were de-stemmed, baked at 40 °C for 2 h, ground, and passed through a 380 μm (40 mesh) sieve. The 0.250 g powder of each sample was accurately weighed, and mixed with 5.0 mL nitric acid and 2.0 mL hydrogen peroxide. After microwave digestion, the sample was diluted to 50 mL with ultrapure water. The blank sample was prepared within the same procedure. The samples were analyzed by ICP. Each sample was repeated three times.

The standards samples were prepared by serial dilution of the stock solution with 1% HNO₃. The concentrations of each element in the standard sample solution were 5, 10, 50, 100, 500, 1000 and 5000 μg/L, respectively. The standard curve was calculated according to the concentration range of 5-5000 μg/L. The content of each element in the working sample was calculated by the standard curve.

D. Statistical analysis

There are significant differences of various element in the order of magnitude as well as outlier data that exceed the value range. In order to reduce the impact of dimensional and order of magnitude on the analysis results, the original quantitative data was processed using the Z-Score standardization method.

The normalization formula is as follows

$$y_i = \frac{x_i - \bar{x}}{s} \quad (1)$$

Where, \bar{x} is the average value of the results of certain element in all samples. s is the standard deviation. After this linear transformation, the mean value of y_i is 0 and the variance is 1. Principal component analysis was then performed using standardized data. The statistical aim is to obtain a principal component with a cumulative contribution greater than 80%. The statistical software employed was SPSS 16.0.

III. RESULTS AND DISCUSSION

A. Wavelength in ICP-OES

Since there were 16 elements included in study, the detection wavelength of each element should be validated carefully to avoid spectral interference. When two or more elements interfered each other, the false positives of specific elements should happen. Therefore, the wavelengths of 16 elements were selected elaborately, as shown in Table 1.

Table 1. The selected wavelength (λ) of 16 elements

element	λ (nm)	element	λ (nm)
As	193.696	Mn	257.610
Se	196.026	Cr	267.716
Zn	206.200	Mg	285.213
Pb	220.353	V	292.464
Cd	228.802	Ca	317.933
Ba	233.527	Cu	327.293
Fe	238.284	Al	396.153
B	249.677	K	766.49

The peak width is also an important parameter. For example, if the peak width of Zn is greater than 0.1 nm, it will interfere with other elements. In order to accurately quantify Zn, its wavelength range was set between 206.18 and 206.22 nm (Figure 1).

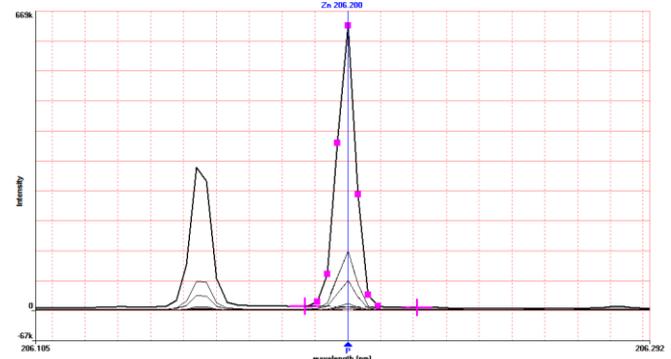


Figure 1. Emission spectra of element Zn at different concentrations

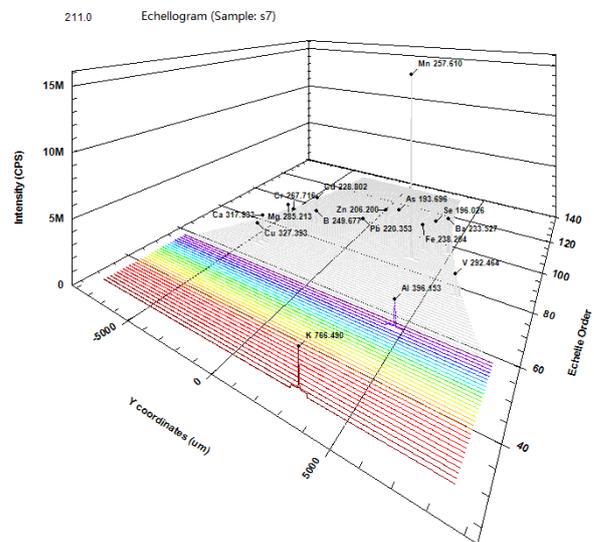


Figure 2. Grating spectrum of mixed standard sample of 16 elements.

A typical grating spectrum of mixed standard sample of all

elements was shown in Figure 2. It can be found that from As at 193.696 nm to K at 766.490 nm, the detection wavelengths of the 16 elements do not interfere with each other.

B. Determination of 16 elements in tobacco leaves

A series of mixed standard solutions at 7 concentration levels were analyzed by ICP method. The standard curve of each element was established with the corresponding peak area. The calculated standard curve equations are listed in Table 2.

Table 2. The standard curves of 16 elements (µg/L)

element	Cali. equation	Corr. coefficient (R ²)
As	y=8.98x+46.1	0.9999
Se	y=6.567x+14.4	0.9999
Zn	y=271.3x+1975.3	0.9999
Pb	y=45.33x+333.7	0.9999
Cd	y=634.7x+2966.8	0.9999
Ba	y=1285x+3796.3	0.9999
Fe	y=933x+4671.3	0.9999
B	y=335.8x+939.6	0.9999
Mn	y=7132x-30392.6	0.9999
Cr	y=767.8x+1850.6	0.9999
Mg	y=1344x-3988.2	0.9999
V	y=447.3x+884.3	0.9999
Ca	y=1142x+38871.2	0.9998
Cu	y=863.5x+1367	0.9999
Al	y=843.5x-13332.4	0.9998
K	y=1106x-59100.8	0.9989

The content of each element in 36 test samples was calculated by substituting the peak area into the standard curve equation. The content of each element in dried tobacco leaves was further calculated according to the sample pretreatment process.

The results showed that the content of Cd, Pb, As, Cr and other heavy metals in all samples did not exceed the limit of China national regulation, limits of pollutants in food (GB2762-2017). Nevertheless, the major elements, trace elements and heavy metals in tobacco leaves from different areas varied obviously. For example, the contents of Pb and Cd in tobacco leaves in Yunnan and Guizhou province are higher than other regions. The possible reason is that the soil in these two areas is more seriously polluted by heavy metals than other areas due to the mining of rare earth metals. As for the tobacco leaves from Peru, the Pb was not detected. But its content of Cr is the highest among all samples.

The quality of tobacco leaves from different producing areas varies greatly. It is difficult to track the source of tobacco leaves relying on the content of several specific elements. Origin traceability of tobacco leaves by multi-elements analysis is necessary to ensure the quality of tobacco leaves.

C. PCA and Clustering Research

Principal component factor (PCA) analysis was performed based on 16 elements. The results show that even by accumulating 5 principal components can the contribution rate only reach 80.9%. The contribution rates of the top five

principal components are listed in Table 3. It indicates that the studied 16 elements are not well correlated. The correlation analysis also indicated that the co-absorption distribution of the 16 elements is not significant. The difference of elements in tobacco leaves was mainly due to the difference of environmental soil rather than the result of tobacco secondary metabolism.

Table 3. Contribution rates of the top five principal components

Component	Initial Eigenvalues	Rate (%)	Cumulative (%)
1	5.075	31.72	31.72
2	3.418	21.362	53.083
3	2.223	13.892	66.974
4	1.227	7.669	74.643
5	1	6.253	80.896

The cumulative contribution rate of the top three principal components can explain 66.97% of the total variation. This result can serve as an effective reference for clustering analysis. A 3D scatter distribution map (Figure 4.) was plotted based on the score vectors of the top three principal components of 36 samples.

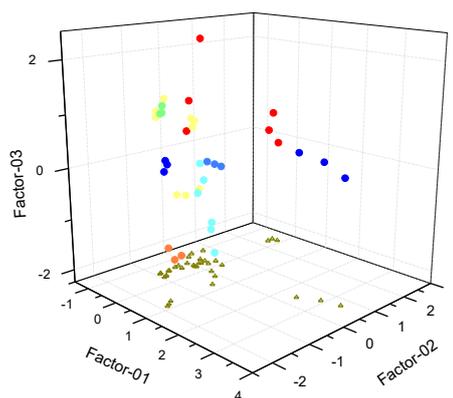


Figure 4. Scatter plots of tobacco leaf samples from different regions based on the scores of the first three principal components in PCA results

In Figure 4, 36 samples can be approximately divided into 10 groups (distinguished by different colors). The results of cluster analysis showed those tobacco leaves from the same area were close and concentrated. The effect of clustering is satisfactory, and there are no outliers.

A clustering analysis was also carried out based on two top principal components. The 2D scatter distribution map was plotted in Figure 4 as the gray-brown projection points. This 2D distribution was based on vector score of factor-01 (principal component 1) and factor-02 (principal component 2) of PCA results. It can be seen that if only the first two principal components are used, 36 samples cannot obtain clear clustering boundaries.

The results indicate that the model established in this study has extracted principal component successfully. The top three principal components achieved good clustering analysis. This mathematical model can serve as a traceability tool for tobacco leaf products based on elemental determination.

IV. CONCLUSION

An ICP-AES method was developed for determination of 16 elements (As, Se, Zn, Pb, Cd, Ba, Fe, B, Mn, Cr, Mg, V, Ca, Cu, Al, K) in tobacco leaves. The wavelengths of 16 elements were selected elaborately to avoid interference. The peak width of each elements was also investigated for the accurate quantification. The standard curves were fitted with a series of mixed standard solutions at 7 concentration levels. All elements calibration equation showed good correlation ($R^2 > 0.9998$). Collected 36 samples from Peru, Zimbabwe, and China were analyzed. The results showed that the content of Cd, Pb, As, Cr and other heavy metals in all samples did not exceed the limit of China national regulation. Principal component factor (PCA) analysis indicated that the studied 16 elements are not well correlated. The difference of elements in tobacco leaves was mainly due to the difference of environmental soil rather than the result of tobacco secondary metabolism. The cumulative contribution rate of the top three principal components can explain 66.97% of the total variation. The PCA result can serve as an effective reference for clustering analysis. 36 samples can be approximately divided into 10 groups. The effect of clustering is satisfactory. The top three principal components achieved good clustering analysis. This mathematical model can serve as a traceability tool for tobacco leaf products. This method may facilitate the tobacco selection and quality identification.

ACKNOWLEDGMENT

This work was supported by Open Project of Jiangsu Key Laboratory for Bio-resources of Saline Soils (No: JKLBS201900X) and Science and Technology Plan Project of Jiangsu Province (No: BY2021462).

REFERENCES

- [1] Malone, R. E., Grundy, Q., Bero, L. A., Tobacco industry denormalisation as a tobacco control intervention: a review. *Tob. Control* 2012, 21, 162–170.
- [2] Banožić, M., Banjari, I., Jakovljević, M., Šubarić, D., Tomas, S., Babić, J., Jokić, S., Optimization of Ultrasound-Assisted Extraction of Some Bioactive Compounds from Tobacco Waste. *Molecules* 2019, 24, 1611.
- [3] Hoffmann, D., Hecht, S. S., 1990, pp. 63–102.
- [4] Kazi, T. G., Jalbani, N., Arain, M. B., Jamali, M. K., Afridi, H. I., Sarfraz, R. A., Shah, A. Q., Toxic metals distribution in different components of Pakistani and imported cigarettes by electrothermal atomic absorption spectrometer. *J. Hazard. Mater.* 2009, 163, 302–307.
- [5] Hong, L., Wang, W., Su, Y., Zhang, G., Su, Y., Zhang, C., Chen, J., Zhe, W., Liu, Z., Cui, J., Mao, D., Wang, J., Identifying the Geographical Origin of Tobacco Leaf by Strontium and Lead Isotopic with Mineral Elemental Fingerprint. *Int. J. Chem. Eng.* 2022, 2022, 1–10.
- [6] Zapranova, P., Dospatliev, L., Angelova, V., Ivanov, K., Correlation between soil characteristics and lead and cadmium content in the aboveground biomass of Virginia tobacco. *Environ. Monit. Assess.* 2010, 163, 253–261.
- [7] Jung, M. C., Thornton, I., Heavy metal contamination of soils and plants in the vicinity of a lead-zinc mine, Korea. *Appl. Geochemistry* 1996, 11, 53–59.
- [8] Liu, H., Zhang, Y., Zhou, X., You, X., Shi, Y., Xu, J., Source identification and spatial distribution of heavy metals in tobacco-growing soils in Shandong province of China with multivariate and geostatistical analysis. *Environ. Sci. Pollut. Res.* 2017, 24, 5964–5975.

- [9] Proshad, R., Zhang, D., Uddin, M., Wu, Y., Presence of cadmium and lead in tobacco and soil with ecological and human health risks in Sichuan province, China. *Environ. Sci. Pollut. Res.* 2020, 27, 18355–18370.
- [10] Wuana, R. A., Okieimen, F. E., Heavy Metals in Contaminated Soils: A Review of Sources, Chemistry, Risks and Best Available Strategies for Remediation. *ISRN Ecol.* 2011, 2011, 1–20.
- [11] Anyidoho, E. K., Teye, E., Agbemaflle, R., Nondestructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm. *Anal. Methods* 2020, 12, 4150–4158.
- [12] Zhao, H., Zhang, G., Wang, D., Liu, Z., Chen, R., Zhang, W., Li, C., Tracing the geographic origin of velvet antlers in China via stable isotope analyses. *RSC Adv.* 2022, 12, 17527–17535.
- [13] Zheng, C., Yang, S., Huang, D., Mao, D., Chen, J., Zhang, C., Kong, W., Liu, X., Xu, Y., Wu, Y., Li, Z., Wang, J., Ye, Y., GC-MS combined with multivariate analysis for the determination of the geographical origin of *Elsholtzia rugulosa* Hemsl. in Yunnan province. *RSC Adv.* 2022, 12, 21287–21296.
- [14] Xu, S., Zhao, C., Deng, X., Zhang, R., Qu, L., Wang, M., Ren, S., Wu, H., Yue, Z., Niu, B., Determining the geographical origin of milk by multivariate analysis based on stable isotope ratios, elements and fatty acids. *Anal. Methods* 2021, 13, 2537–2548.
- [15] Han, C., Dong, S., Li, L., Wei, F., Zhou, Y., Gao, Q., The effect of the seasons on geographical traceability of salmonid based on multi-element analysis. *Food Control* 2020, 109, 106893.
- [16] Ricardo, F., Mamede, R., Bispo, R., Santos, A., Ferreira da Silva, E., Patinha, C., Calado, R., Cost-efficiency improvement of bivalves shells preparation when tracing their geographic origin through ICP-MS analysis of elemental fingerprints. *Food Control* 2020, 118, 107383.
- [17] Chung, I.-M., Kim, J.-K., Lee, J.-K., Kim, S.-H., Discrimination of geographical origin of rice (*Oryza sativa* L.) by multielement analysis using inductively coupled plasma atomic emission spectroscopy and multivariate analysis. *J. Cereal Sci.* 2015, 65, 252–259.
- [18] Cui, D., Liu, Y., Yu, H., Wang, Z., Mao, X., Geographical traceability of soybean based on elemental fingerprinting and multivariate analysis. *J. Consum. Prot. Food Saf.* 2021, 16, 323–331.
- [19] Mahne Opatić, A., Nečemer, M., Budič, B., Lojen, S., Stable isotope analysis of major bioelements, multi-element profiling, and discriminant analysis for geographical origins of organically grown potato. *J. Food Compos. Anal.* 2018, 71, 17–24.
- [20] Anderson, K. A., Magnuson, B. A., Tschirgi, M. L., Smith, B., Determining the Geographic Origin of Potatoes with Trace Metal Analysis Using Statistical and Neural Network Classifiers. *J. Agric. Food Chem.* 1999, 47, 1568–1575.
- [21] Chaguri, M. P., Maulvault, A. L., Nunes, M. L., Santiago, D. A., Denadai, J. C., Fogaça, F. H., Sant’Ana, L. S., Ducatti, C., Bandarra, N., Carvalho, M. L., Marques, A., Different tools to trace geographic origin and seasonality of croaker (*Micropogonias furnieri*). *LWT - Food Sci. Technol.* 2015, 61, 194–200.
- [22] Kang, X., Zhao, Y., Shang, D., Zhai, Y., Ning, J., Ding, H., Sheng, X., Identification of the geographical origins of sea cucumbers in China: The application of stable isotope ratios and compositions of C, N, O and H. *Food Control* 2020, 111, 107036.
- [23] Lin, T., Chen, X., Li, B., Chen, P., Guo, M., Zhou, X., Zhong, S., Cheng, X., Geographical origin identification of *Spodoptera litura* (Lepidoptera: Noctuidae) based on trace element profiles using tobacco as intermedium planted on soils from five different regions. *Microchem. J.* 2019, 146, 49–55.
- [24] Xiang, B., Cheng, C., Xia, J., Tang, L., Mu, J., Bi, Y., Simultaneous identification of geographical origin and grade of flue-cured tobacco using NIR spectroscopy. *Vib. Spectrosc.* 2020, 111, 103182.
- [25] Zhao, J., Hu, C., Zeng, J., Zhao, Y., Zhang, J., Chang, Y., Li, L., Zhao, C., Lu, X., Xu, G., Study of polar metabolites in tobacco from different geographical origins by using capillary electrophoresis–mass spectrometry. *Metabolomics* 2014, 10, 805–815.
- [26] Zhang, L., Wang, X., Guo, J., Xia, Q., Zhao, G., Zhou, H., Xie, F., Metabolic Profiling of Chinese Tobacco Leaf of Different Geographical Origins by GC-MS. *J. Agric. Food Chem.* 2013, 61, 2597–2605.