Discriminating the Geographical Origin of Soybeans Using ICP-MS and Chemometrics

Elemental profiling of soybeans using Agilent 7850 ICP-MS and Mass Profiler Professional software

Authors

Maria del Mar Aparicio Muriana[a], Yunhe Hong[a], Cynthia Adaku Chilaka[a], Brian Quinn[a], Alfredo M. Montes-Niño[b], Nicholas Birse[a], Christopher T. Elliott[a,c], Glenn D. Woods[d], M. Lorna De Leoz[d]

[a] National Measurement Laboratory: Centre of Excellence in Agriculture and Food Integrity, Queen's University Belfast, UK

[b] Microbioticos Paraguay SRL, San Lorenzo, Paraguay

[c] School of Food Science and Technology, Thammasat University, Pathum Thani, Thailand

[d] Agilent Technologies, Inc.

Introduction

Soybean *(Glycine max)* is one of the world's most important agricultural commodities, serving as a primary source of protein for both human consumption and animal feed. As global demand for soybeans increases, so do concerns about its environmental and regulatory implications. Soybean production is closely linked to deforestation and land-use change, particularly in South America, where large-scale agricultural expansion threatens natural ecosystems. To ensure sustainable supply chains and prevent links to illegal deforestation or environmental degradation, regulatory restrictions and certification schemes, such as the 2023 EU Deforestation Regulation, are being implemented.

Geographical origin verification is essential for regulatory compliance and ensuring consumer trust in sustainable sourcing claims. It enables buyers to verify that soybeans have been harvested in regions that comply with environmental standards, are free from deforestation, and adhere to fair agricultural practices.

For producers, origin verification protects the value of their products and demonstrates accountability in supply chain management. However, distinguishing the origin of soybeans based on traditional methods, such as documentation or certification alone, can be challenging due to the complexity of global supply chains. In addition, mislabeling of geographical origin is one of the most common forms of fraud in the food industry, underscoring the need for advanced methods to verify origin and enhance traceability.

The elemental composition of soybean reflects its growing environment, including soil properties, climate conditions, and agricultural practices. This elemental fingerprint provides a powerful means of determining geographical origin of the crop. Following microwave-assisted acid digestion of the samples,[1] ICP-MS combined with advanced data analysis tools enables precise multi-elemental profiling of soybean samples, reflecting their geographical origin. This approach has been successfully applied to other food products to distinguish geographical origin and verify authenticity.[2–4]

Agilent ICP-MS instruments incorporate an advanced Octopole Reaction System (ORS^4) designed to eliminate common polyatomic interferences using helium (He) collision mode combined with kinetic energy discrimination (KED). He KED mode enables standardized cell settings across multiple elements, delivering high-quality, multi-element data sets essential for studies involving food authenticity and origin discrimination. Additionally, the Agilent 7850 ICP-MS system features a linear dynamic range over 10 or 11 orders of magnitude, enabling the simultaneous measurement of both major and trace elements within the same analytical run.

For data processing, Agilent Mass Profiler Professional (MPP) software provides advanced chemometric tools to perform statistical analysis and visualizations of large-scale mass spectrometry data from both Agilent and non-Agilent data. MPP offers a variety of data classification methods, allowing users to develop reliable models for predicting attributes such as the geographical origin of samples.

In this study, 330 soybean samples were analyzed using the 7850 ICP-MS. The samples were collected across four growing seasons from seven major producers: Brazil, the USA, Argentina, Paraguay, China, India, and Canada. The elemental data obtained from the 7850 ICP-MS was processed using MPP software to develop chemometric models for geographical origin discrimination.

By applying ICP-MS and advanced chemometric modeling to soybean authentication, this study contributes to improving supply chain transparency, combatting deforestation, and supporting sustainable agricultural practices. Our results demonstrate a scalable and robust analytical approach for verifying soybean origin, promoting sustainability, and ensuring compliance with environmental regulations.

Experimental

Calibration standards

Calibration was performed using two separate multipoint external calibration curves, tailored to the elemental concentration range.

1. For Na, Mg, Ca, Fe, and K, calibration was prepared using the Agilent environmental calibration standard (p/n 5183-4688), up to 200 mg/L.
2. For the remaining elements, calibration was prepared using a combination of calibration standard-1 (p/n 8500-6944), calibration standard-2A (8500-6940), and calibration standard-4 (8500-6942), up to 400 µg/L.

A calibration blank was also included in the worklist. All calibration standards were prepared in 2% nitric acid (HNO_3, 67–69%) and 0.5% hydrochloric acid (HCl) in Milli-Q deionized (DI) water. The indium (In) and bismuth (Bi) internal standard (ISTD) solution was prepared from an Agilent single-element In standard (p/n 8500-6946) and Bi standard (p/n 8500-6936) at 100 µg/L. The ISTD solution was automatically mixed with the sample online using a tee connector.

Sample preparation

Various collaborators provided 330 soybean samples from seven countries over four growing seasons. The samples were stored in plastic zip bags before sample preparation and analysis.

Dried soybeans were ground in a kitchen processor. A 0.100 g ± 0.005 g portion of the sample was accurately weighed into a 50-mL polytetrafluoroethylene (PTFE) tube. Each sample was then predigested with 2 mL of HNO_3 (67–69%) under a chemical hood for at least 15 hours, with the lid loosely tightened.

Reagent blanks were prepared using the same procedure but without adding a sample (i.e., only 2 mL of HNO_3). A reference material (RM), consisting of an organic soybean meal analytical standard obtained from Elemental Microanalysis (Okehampton, UK), was treated identically to the samples.

After pre-digestion, 2 mL of hydrogen peroxide (H_2O_2, 30%) was added to each tube (samples, reagent blanks, and RM), and the following microwave digestion program was applied: the samples were heated from room temperature to 54 °C in 5 min, held at 54 °C for 15 min, then heated to 65 °C over the next 5 min, and held at 65 °C for 10 min. Finally, the samples were heated to 95 °C and kept at that temperature for 30 min. After cooling, the tubes were made up to 20 g with DI water containing 0.5% HCl.

All samples, reagent blanks, and RM were prepared using the same procedure (Figure 1).

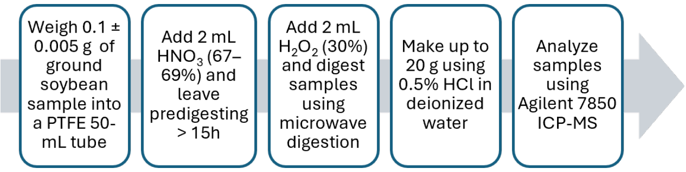

Figure 1. Analytical workflow used for the analysis of soybean by ICP-MS.

Instrumentation

The Agilent 7850 ICP-MS with Agilent SPS 4 autosampler was used for the analysis. The 7850 was configured with the standard MicroMist nebulizer, quartz spray chamber, 2.5 mm injector quartz torch, and nickel cones. All elements were measured using He collision cell mode except for B, which was measured using no gas mode. The 7850 ICP-MS was controlled using Agilent ICP-MS MassHunter 5.1 software, and the data was processed using MPP software (version 15.1).

Table 1. Agilent 7850 ICP-MS operating conditions.

ICP-MS Parameters	No Gas Mode	He Mode
Plasma Mode	General purpose	
RF Power (W)	1550	
RF Matching (V)	0.80	
Sampling Depth (mm)	10	
Nebulizer Gas Flow (L/min)	1.07	
Spray Chamber Temperature (°C)	2.0	
Cell Gas Flow Rate (L/min)	0.0	0.5
Dilution Gas Flow Rate (mL/min)	0.0	
Lens Tune	Autotune	
Energy Discrimination (V)	5	
Number of Elements	1	39

Results and discussion

Detection limits

During method development, 40 elements were measured and used in the class-prediction modeling. Calibration coefficients (R) resulting from the analysis were higher than 0.9995 for all elements (Figure 2). Instrument detection limits (IDLs) were calculated as 3 x standard deviation (SD) of the concentration of calibration blanks (n=10). Method detection limits (MDLs) were calculated after applying a 200x dilution factor (0.100 g ± 0.005 g of soybean digested and diluted to a final weight of 20 g). The IDLs and MDLs are shown in Table 2. The low IDLs demonstrate the suitability of the 7850 ICP-MS for the determination of trace elements.

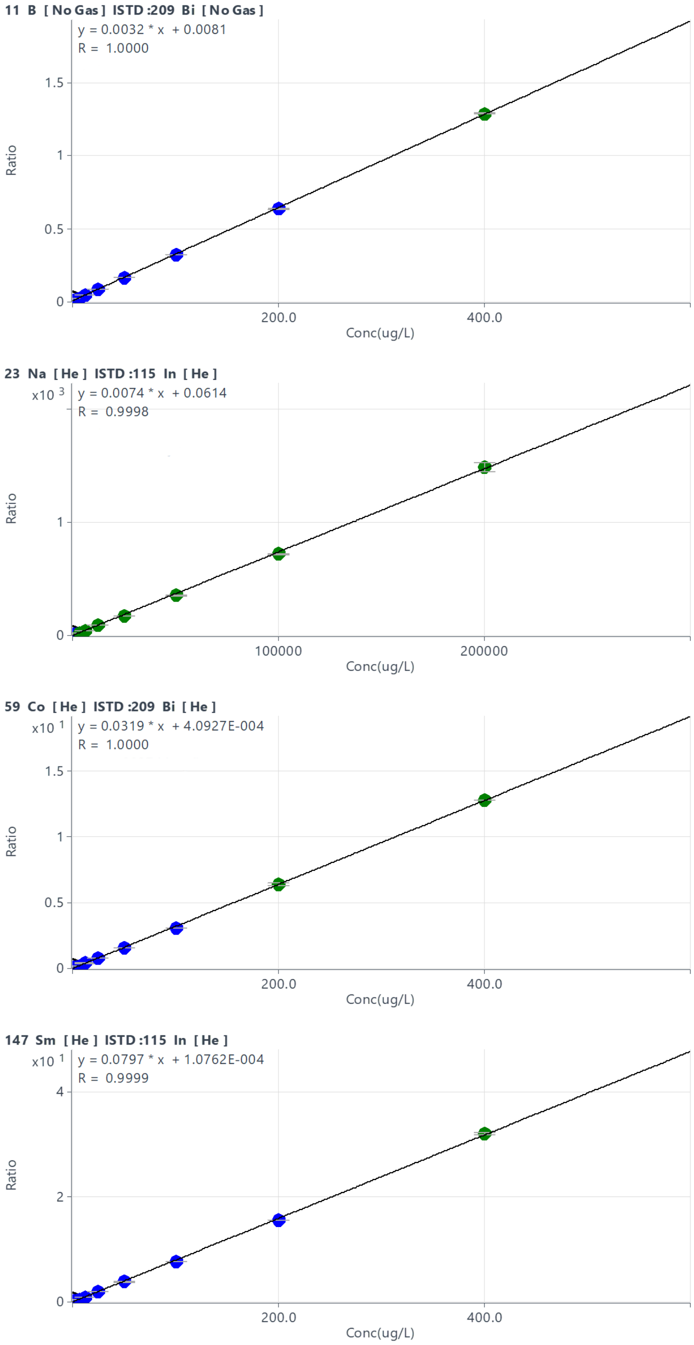

Figure 2. Example of calibration curves for B, Na, Co, and Sm determined following the described operating procedure.

Table 2. Cell mode conditions, internal standard, instrumental detection limits, and method detection limits of the 40 elements used to discriminate the geographical origin of soybeans.

Element	Mode	ISTD	IDL (µg/kg)	MDL (µg/kg)
11 B	No gas	209 Bi	0.325	64.99
23 Na	He	115 In	0.798	159.60
24 Mg	He	209 Bi	0.278	55.51
27 Al	He	115 In	0.314	62.86
39 K	He	115 In	2.481	496.3
43 Ca	He	115 In	11.21	2242
47 Ti	He	209 Bi	0.261	52.22
51 V	He	209 Bi	0.009	1.821
52 Cr	He	209 Bi	0.006	1.247
55 Mn	He	209 Bi	0.024	4.850
56 Fe	He	209 Bi	0.147	29.36
59 Co	He	209 Bi	0.003	0.597
60 Ni	He	209 Bi	0.029	5.836
63 Cu	He	209 Bi	0.024	4.825
66 Zn	He	209 Bi	0.131	26.18
71 Ga	He	209 Bi	0.006	1.207
72 Ge	He	209 Bi	0.015	2.996
78 Se	He	209 Bi	0.088	17.51
85 Rb	He	115 In	0.010	1.943
88 Sr	He	115 In	0.006	1.103
89 Y	He	115 In	0.002	0.341
95 Mo	He	209 Bi	0.004	0.738
111 Cd	He	209 Bi	0.002	0.494
133 Cs	He	115 In	0.002	0.310
135 Ba	He	115 In	0.014	2.731
139 La	He	115 In	0.002	0.310
140 Ce	He	115 In	0.001	0.290
141 Pr	He	115 In	0.001	0.190
146 Nd	He	115 In	0.002	0.310
147 Sm	He	115 In	0.001	0.253
153 Eu	He	115 In	0.002	0.316
157 Gd	He	115 In	0.002	0.316
159 Tb	He	115 In	0.001	0.290
163 Dy	He	115 In	0.002	0.310
165 Ho	He	115 In	0.001	0.290
166 Er	He	115 In	0.001	0.290
169 Tm	He	115 In	0.001	0.290
172 Yb	He	115 In	0.001	0.190
175 Lu	He	115 In	0.001	0.253
185 Re	He	209 Bi	0.001	0.253

Internal standard stability

Indium and bismuth were used as ISTDs as they were not expected to be present in the soybean samples. The analytical sequence, consisting of 330 soybean sample digests prepared in triplicate, method blanks, and RM control checks, was analyzed over several consecutive days.

As shown in Figure 3, the ISTD recovery measurements were within the ± 25% limits (indicated by the red dotted lines) for all of the samples. The recovery test shows that the plasma of the 7850 ICP-MS was able to decompose the variable sample matrices effectively, enabling excellent stability to be maintained over a 12-hour run. Also, no significant matrix deposition occurred on the interface during the sequence.

Reference Material check

To verify measurement accuracy, two replicates of the soybean meal organic analytical standard were analyzed at the start and end of every sample batch. Elemental recoveries were calculated for Cu, Al, K, Fe, Ca, Mg, Na, B, Zn, and Mn, as manufacturer-reported concentrations were available for these elements. The recoveries for all monitored elements were within 100 ± 10%, except for B and Al, which were within 100 ± 30%.

The robustness of the elemental measurements was further evaluated by comparing the concentrations of all monitored elements detected in the RM at the start and the end of each sample batch, regardless of whether the true RM concentration was available from the manufacturer. These concentration checks were performed considering all RM injections across different batches and on several days when the samples were analyzed. After analyzing more than 100–120 samples per batch, the percentage difference between the concentrations measured was less than 14.2% for all elements included in the model, with an average percentage difference of 4.5%, demonstrating the consistency and reliability of the method over time (Table 3).

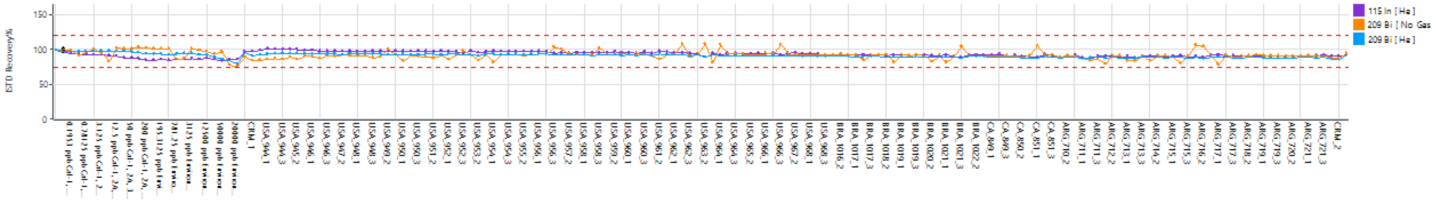

Figure 3. Recovery of In and Bi ISTDs in no gas and He mode in one run (172 analyses) over a 12-hour sequence using the Agilent 7850 ICP-MS.

Table 3. QC check solution measured before and after the six-hour run using the Agilent 7850 ICP-MS. Concentration units: µg/L (ppb).

Element	Variation (%)
11 B	8.8
23 Na	1.4
24 Mg	2.3
27 Al	4.9
39 K	2.9
43 Ca	0.9
47 Ti	7.7
51 V	0.8
52 Cr	1.5
55 Mn	0.3
56 Fe	0.5
59 Co	2.1
60 Ni	2.8
63 Cu	4.6
66 Zn	0.6
71 Ga	2.5
72 Ge	13.7
78 Se	10.6
85 Rb	0.4
88 Sr	0.9

Element	Variation (%)
89 Y	1.3
95 Mo	0.6
111 Cd	5.4
133 Cs	2.5
135 Ba	1
139 La	0.6
140 Ce	2.4
141 Pr	0.2
146 Nd	2
147 Sm	0.5
153 Eu	10
157 Gd	0.2
159 Tb	13.1
163 Dy	4.8
165 Ho	14.1
166 Er	8.4
169 Tm	11.5
172 Yb	5.7
175 Lu	14.2
185 Re	10.5

Spike recoveries and evaluation of matrix effects

A spike recovery test was performed on a soybean sample from Canada to evaluate the method's accuracy and matrix effects. Three replicates of the spiked sample were analyzed, and the recovery percentages are presented in Table 4. The average of the spike recoveries (n=3) fell within 100 ± 18%, demonstrating that matrix effects from the sample did not significantly impact the measured elements.

Table 4. Elemental recovery results for a spiked Canadian soybean sample, prepared in triplicate and analyzed by Agilent 7850 ICP-MS.

Element	Unspiked Sample Concentration (µg/kg) (Mean, n=3)	Spike Concentration (µg/kg)	Spiked Sample Concentration (µg/kg) (Mean, n=3)	Recovery (%) (Mean, n=3)
11 B	25253	10000	30891	87.6
23 Na	4009	10000	12104	86.4
24 Mg	2339869	1000000	3151097	94.3
27 Al	11795	1000	15150	118.4
39 K	1735607	1000000	17657288	96.2
43 Ca	2069200	1000000	2839021	92.5
47 Ti	623.7	1000	1659	102.2
51 V	27.97	1000	1196	116.4
52 Cr	293.3	1000	1501	116.1
55 Mn	24564	10000	35898	103.9
56 Fe	80979	10000	94590	104.0
59 Co	65.35	100.0	186.2	112.6
60 Ni	1201	100.0	1292	99.3
63 Cu	12546	1000	13624	100.6
66 Zn	38666	1000	39036	98.4
71 Ga	3.45	10.00	13.21	98.3
72 Ge	(0.40) <MDL	10.00	10.81	103.9
78 Se	86.48	10.00	92.74	96.1
85 Rb	7743	1000	8755	100.1
88 Sr	2680	1000	3618	98.3
89 Y	3.92	10.00	14.51	104.2
95 Mo	1851	100.0	1969	100.9
111 Cd	40.45	10.00	48.52	96.2
133 Cs	8.87	10.00	18.97	100.6
135 Ba	1104	100.0	1217	101.0
139 La	7.93	1.00	8.73	97.7
140 Ce	17.06	1.00	19.23	106.5
141 Pr	1.75	1.00	2.92	106.3
146 Nd	7.69	1.00	8.63	99.3
147 Sm	1.46	1.00	2.12	86.3
153 Eu	0.36	1.00	1.35	99.8
157 Gd	1.21	1.00	2.44	110.5
159 Tb	(0.19) <MDL	1.00	1.27	106.7
163 Dy	0.99	1.00	2.04	102.4
165 Ho	(0.23) <MDL	1.00	1.34	108.7
166 Er	0.51	1.00	1.67	110.2
169 Tm	(0.06) <MDL	1.00	1.13	106.9
172 Yb	0.34	1.00	1.41	105.3
175 Lu	(0.14) <MDL	1.00	1.28	112.1
185 Re	(0.00) <MDL	1.00	0.95	95.5

Data analysis using MPP software

The ICP-MS data for 40 elements measured in 330 soybean samples from seven countries were combined and imported into the MPP chemometric software for statistical analysis. The MPP software provides a range of statistical and visualization tools, including t-test, analysis of variance (ANOVA), model-building algorithms, Box and Whisker plots, correlation, and clustering analysis.

Principal component analysis

A Principal Component Analysis (PCA), with a P cut-off value <0.05, was performed to determine whether there were significant differences between soybean sample groups based on geographical origin. The PCA technique evaluates the relative contribution of elements to the separation of the groups.

The 3D PCA scores plot (Figure 4) revealed variation between soybean samples from different countries based on elemental profiling. A total of 50.97% of the variance ratio was captured in the first three dimensions: PC1 (x-axis, 32.41%), PC2 (y-axis, 10.68%), and PC3 (z-axis, 7.88%).

Based on the PCA scores and loading plots for PC1 vs PC2 (Figure 5) and PC1 vs PC3 (Figure 6), the primary elements contributing to geographical separation were Nd, Sm, and Gd for PC1; B, Co, and Zn for PC2; and Ni, Cr, and Cd for PC3.

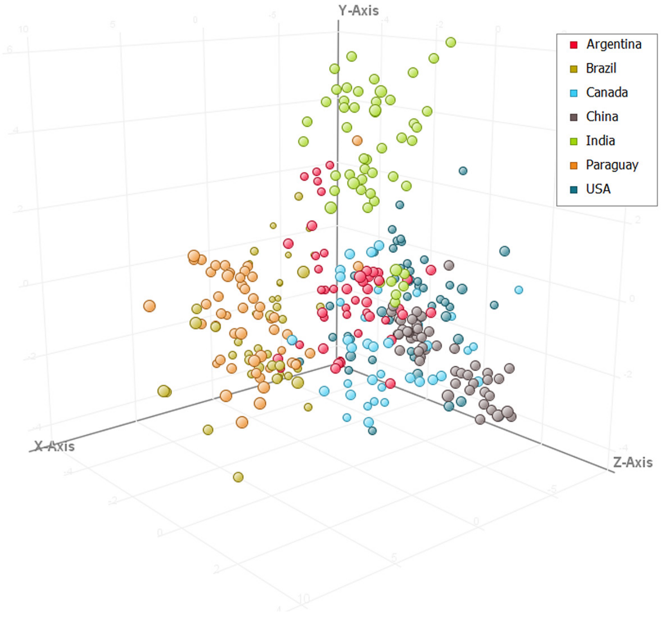

Figure 4. 3D-PCA scores plot of 330 soybean samples. The plot's x-axis, y-axis, and z-axis represent the three major principal components of the PCA results (PC1, PC2, and PC3).

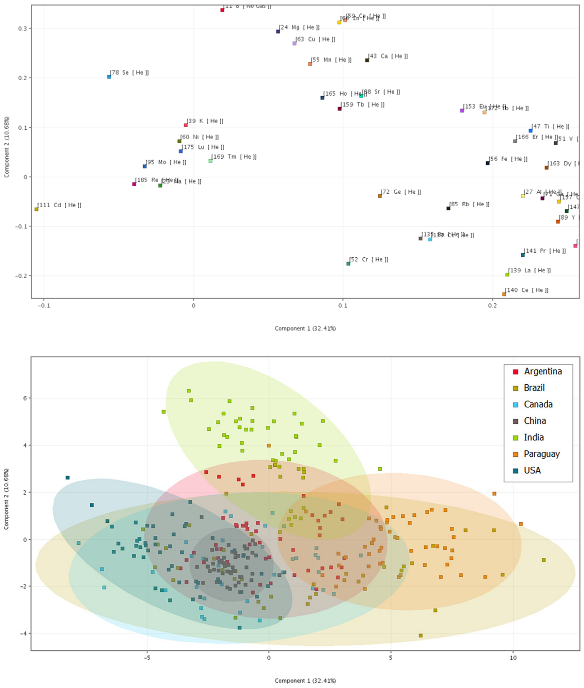

Figure 5. PCA loadings (top) and PCA scores plot (bottom) of 330 soybean samples showing the discrimination in geographical origin based on elemental profiling. The x-axis and y-axis of the plot represent the two major principal components of the PCA results (PC1 and PC2).

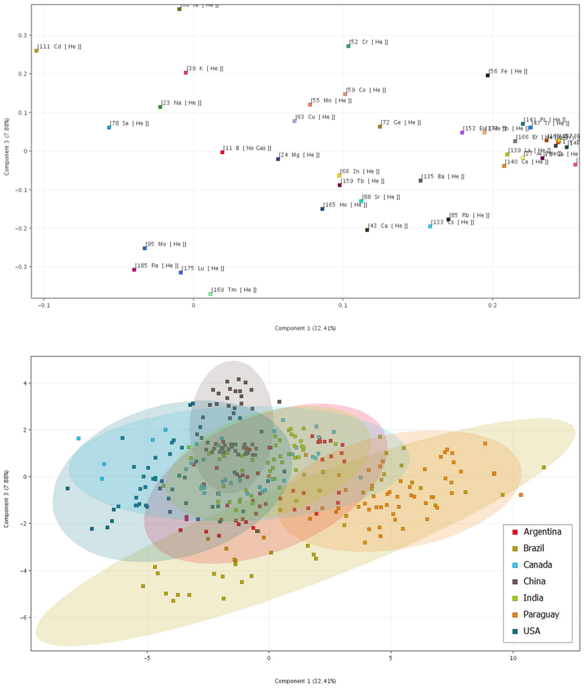

Figure 6. PCA loadings (top) and PCA scores plot (bottom) of 330 soybean samples showing the discrimination in geographical origin based on elemental profiling. The x-axis and y-axis of the plot represent the first and third major principal components of the PCA results (PC1 and PC3).

Class prediction analysis

Class prediction analysis is a powerful technique for objectively assigning unknown samples to predefined groups. The MPP software offers multiple class prediction algorithms. For this study, Linear Discriminant Analysis (LDA) was selected to develop prediction models for identifying the geographical origin of soybean samples based on their elemental composition. The dataset was randomly divided into a training set (80%) and a test set (20%), ensuring a balanced representation across the seven countries.

The training set was used to develop the model by identifying the linear combinations of elements that best separate the predefined geographic groups. The LDA scores plot was used to assess sample separation and clustering (Figure 7) visually. The test set, consisting of 66 unknown samples, was then used to evaluate the model's classification accuracy. The predicted and actual origin of the test set are shown in Table 5. The origin of 64 of the 66 samples was correctly assigned using the LDA model, with confidence measures ranging from 0.226 to 0.947. This equates to a 98.5% correct prediction of the geographical location, indicating the model's effectiveness and reliability.

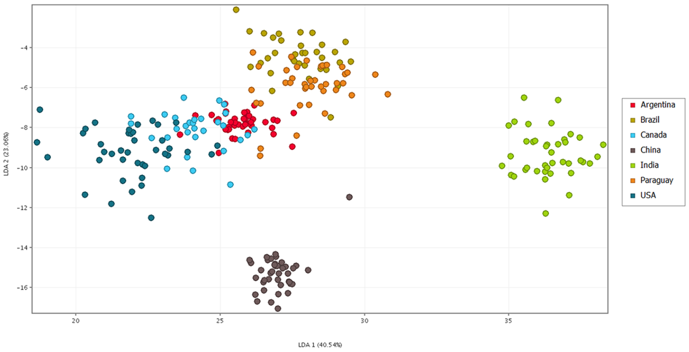

Figure 7. LDA scores plot for the training set.

Table 5. Results of predicted locations from the 66 soybean test samples.

Sample Name	Actual Location	LDA Prediction Model	
		Predicted Location	Confidence Measure
Unknown A1	Argentina	Argentina	0.726
Unknown A2	Argentina	Argentina	0.776
Unknown A3	Argentina	Argentina	0.148
Unknown A4	Argentina	Argentina	0.330
Unknown A5	Argentina	Canada	0.052
Unknown A6	Argentina	Argentina	0.688
Unknown A7	Argentina	Argentina	0.660
Unknown A8	Argentina	Argentina	0.750
Unknown A9	Argentina	Argentina	0.734
Unknown A10	Argentina	Argentina	0.773
Unknown B1	Brazil	Brazil	0.387
Unknown B2	Brazil	Brazil	0.521
Unknown B3	Brazil	Brazil	0.619
Unknown B4	Brazil	Brazil	0.411
Unknown B5	Brazil	Brazil	0.527
Unknown B6	Brazil	Brazil	0.226
Unknown B7	Brazil	Brazil	0.373
Unknown B8	Brazil	Brazil	0.416
Unknown Ca1	Canada	Canada	0.469
Unknown Ca2	Canada	Canada	0.700
Unknown Ca3	Canada	Canada	0.438
Unknown Ca4	Canada	Canada	0.686
Unknown Ca5	Canada	Canada	0.259
Unknown Ca6	Canada	Canada	0.542
Unknown Ca7	Canada	Canada	0.807
Unknown Ca8	Canada	Canada	0.771
Unknown Ch1	China	China	0.858
Unknown Ch2	China	China	0.848
Unknown Ch3	China	China	0.887
Unknown Ch4	China	China	0.947
Unknown Ch5	China	China	0.814
Unknown Ch6	China	China	0.917
Unknown Ch7	China	China	0.829
Unknown Ch8	China	China	0.940
Unknown Ch9	China	China	0.838
Unknown Ch10	China	China	0.619

Table 5 continued. Results of predicted locations from the 66 soybean test samples.

Sample Name	Actual Location	LDA Prediction Model	
		Predicted Location	Confidence Measure
Unknown I1	India	India	0.681
Unknown I2	India	India	0.681
Unknown I3	India	India	0.811
Unknown I4	India	India	0.682
Unknown I5	India	India	0.692
Unknown I6	India	India	0.845
Unknown I7	India	India	0.726
Unknown I8	India	India	0.735
Unknown I9	India	India	0.812
Unknown I10	India	India	0.777
Unknown P1	Paraguay	Paraguay	0.637
Unknown P2	Paraguay	Paraguay	0.717
Unknown P3	Paraguay	Paraguay	0.619
Unknown P4	Paraguay	Paraguay	0.678
Unknown P5	Paraguay	Paraguay	0.785
Unknown P6	Paraguay	Paraguay	0.330
Unknown P7	Paraguay	Paraguay	0.615
Unknown P8	Paraguay	Paraguay	0.395
Unknown P9	Paraguay	Paraguay	0.626
Unknown P10	Paraguay	Paraguay	0.633
Unknown U1	USA	USA	0.249
Unknown U2	USA	USA	0.512
Unknown U3	USA	Canada	0.179
Unknown U4	USA	USA	0.367
Unknown U5	USA	USA	0.593
Unknown U6	USA	USA	0.680
Unknown U7	USA	USA	0.533
Unknown U8	USA	USA	0.533
Unknown U9	USA	USA	0.566
Unknown U10	USA	USA	0.532

Conclusion

This study demonstrated the successful application of ICP-MS and chemometric modeling for the geographical authentication of soybeans. The Agilent 7850 ICP-MS enabled the simultaneous measurement of 40 major and trace elements, while the Agilent Mass Profiler Professional (MPP) software provided robust statistical tools for origin classification. This approach effectively differentiated soybean samples from seven major producing countries by integrating multi-elemental profiling and advanced statistical modeling. When applied to test samples, LDA-based class prediction models achieved high classification accuracy (98.5%). The methodology demonstrated excellent reproducibility, minimal matrix effects, and high measurement stability over multiple analysis days.

These findings highlight the potential of ICP-MS and chemometric modeling as a scalable and reliable solution for verifying soybean origin, ensuring supply chain transparency, combating food fraud, and supporting sustainable agricultural practices. The study further emphasizes the importance of traceability and compliance with environmental regulations, such as the 2023 EU Deforestation Regulation.

This multi-elemental authentication approach can be extended to other agricultural commodities, reinforcing trust in sustainable sourcing claims and enhancing global food security and regulatory compliance.

References

1. Sakai, K.; Takahashi, J.; McCurdy, Ed. Application of the Agilent 7900 ICP-MS with Method Automation function for the routine determination of trace metallic components in food CRMs, Agilent publication, 5991-4556EN
2. Xu, F. et.al. Authenticating Rice by Elemental Profiling Using ICP-MS and Statistical Modeling, Agilent publication, 5994-4043EN
3. Jain, V.; Dong, S.; Kanrar, B. Authenticating Geographical Origin of Tea Using ICP-MS and Agilent Mass Profiler Professional Software, Agilent publication, 5994-4583EN
4. Jain, V; Dong, S; Sen, P.; Kar, Prasenjit. Using ICP-MS Trace Element Fingerprint and MPP to Confirm the Region of Origin of High-Value Indian Teas, Agilent ICP-MS Journal, 91, **2023**, 5994-5585EN

More information

Aparicio-Muriana, M.M., Hong, Y., Chilaka, C.A. et al. Mapping the global origins of soybean: a study using ICP-MS and chemometrics. *npj Sci Food* 9, **2025,** 263, https://doi.org/10.1038/s41538-025-00630-5

Agilent University Relations Thought Leader program and Grant ID #4628

www.agilent.com/chem/7850icpms

DE-006865

This information is subject to change without notice.

© Agilent Technologies, Inc. 2026
Published in the USA, January 12, 2026
5994-8404EN

Agilent

Trusted Answers