Tracing origins of uranium ore concentrates (UOCs) by multidimensional statistics analysis of rare-earth impurities

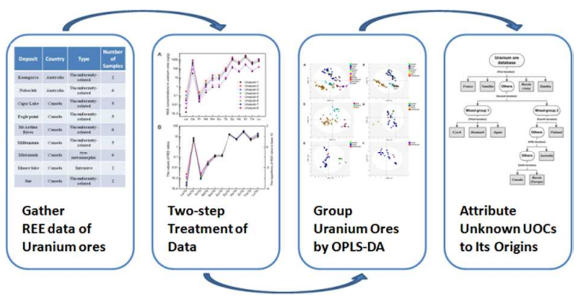
Mu Lin[a,b], Yonggang Zhao[a,*], Lifei Zhao[a], Lili Li[a], Fan Wang[a], Liuchao Zhu[a], Xiaonan Hu[a] and Wei Ning[a]

[a] Department of Radiochemistry, China Institute of Atomic Energy, Beijing, China

[b] State Nuclear Power Research Institute, Beijing, China

* Corresponding author. Box 8, P.O. BOX 275, Beijing, China. Tel.: +86-10-69358448; Fax: +86-10-69357936. E-mail address: zhaoyg@ciae.ac.cn

Graphical abstract



A novel approach using REE data of uranium ores instead of UOC samples to trace the origins of unknown UOCs.

Journal of Analytical Atomic Spectrometry

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

Paper

Tracing origins of uranium ore concentrates (UOCs) by multidimensional statistics analysis of rare-earth impurities

Mu Lin,[ab] Yonggang Zhao,*[a] Lifei Zhao,[a] Lili Li,[a] Fan Wang,[a] Liuchao Zhu,[a] Xiaonan Hu[a] and Wei Ning[a]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
DOI: 10.1039/b000000x

Identifying the origin of uranium ore concentrate (UOC) has become a research hotspot in nuclear forensics. In the present study, a novel approach using REE data of uranium ores instead of UOC samples to trace the origins of unknown UOCs was established with the help of iterative OLPS-DA. In this paper, a database was established by gathering published REE data of uranium ores from different countries. The data were transformed by two-step pretreatment so that they could show stronger geographical and geological indications and the REE patterns of uranium ores could be comparable with each other and those of unknown UOCs. The efficiencies of different multidimensional statistics techniques in pattern recognition, including principal component analysis (PCA), partial least squares discriminant analysis (PLS-DA) and orthogonal partial least squares discriminant analysis (OPLS-DA), were compared and the results showed that OPLS-DA was the most effective method when group uranium ores from different origins. With the help of iterative OPLS-DA, samples from 10 different countries were totally distinguished within several iterations and the discrimination orders formed a decision tree, or attribution strategy. The results indicated that it was feasible to attribute an unknown UOC to its country, deposit and type of uranium ore by comparing the unknown UOC with uranium ores from different origins according to the decision tree. Eventually, the study verifies that REE pattern is a powerful geographical and geological indicator when identifying the origin of UOC by analysing REE data of uranium ores from different origins with the help of iterative OLPS-DA, and iterative OLPS-DA has better potential in nuclear forensics studies comparing to PCA and PLS-DA, which provides support to fighting against illicit trafficking of nuclear materials.

1. Introduction

At present, the illicit trafficking of nuclear materials is a potential threat to world peace and nuclear forensics has become an increasingly important tool to fight against the growing problem.[1] Uranium ore concentrate (UOC), a product of the front end of the nuclear fuel cycle, usually gets involved in the international trading of nuclear materials. UOCs are yielded from uranium ores by leaching and extracting processes. Their physical properties, chemical compositions and especially impurity contents highly depend on the nature of raw uranium ores and suitable milling methods[2] and have geographical and geological signatures to some extent. Therefore, it is feasible to attribute an unknown UOC to its origin (e.g. country, deposit, type of uranium ore and milling methods, etc.) by elemental[3-5], anionic[2-3] and isotopic[5-8] signatures. Attribution of UOCs has become a proof of the ability to fight against the illicit trafficking of nuclear materials and turned into a researching hotspot in nuclear forensics.

Comparing the shapes of the rare-earth element (REE) patterns of UOCs from different countries, it is believed that REE pattern contains geographical information.[5] On the other hand, REE pattern is also regarded as a powerful geological signature in the former studies of identifying the origin of UOCs[3-5], considering that: (1) the REE pattern of uranium ore, especially the LREE/HREE, is influenced by metallogenic conditions;[9] (2) the REE pattern gets hardly changed during milling processes since REEs have similar chemical properties;[10] (3) there are a lot of reliable measurement methods and published data of REEs in uranium ore/product. So the geological and geographical signatures implied in the REE patterns of UOCs became the foundation of the studies of attribution of UOCs.

However, it is not realistic to identify whether some UOCs come from the same origin by comparing dozens of REE concentrates directly, because the degree of the similarity or dissimilarity of different REE patterns is largely influenced by intuitive judgment that varies with each individual. To quantificationally characterize the common feature of REE patterns in UOCs from a certain origin, multidimensional statistics techniques, such as principal component analysis (PCA)[4] and partial least squares discriminant analysis (PLS-DA)[11], have been employed to reduce dimensionality, visualize

Table 1 Description of the uranium ores in the comparison database

Country	Number of samples	Deposit	Location	Deposit Type	Instrument	Reference
Australia	2	Koongarra	Kombolgie Basin	Unconformity-related	SIMS	9
	6	Nabarlek	Kombolgie Basin	Unconformity-related	SIMS	9
Canada	5	Cigar Lake	Athabasca Basin	Unconformity-related	SIMS	9
	5	Eagle point	Athabasca Basin	Unconformity-related	SIMS	9
	6	McArthur River	Athabasca Basin	Unconformity-related	SIMS	9
	5	Millennium	Athabasca Basin	Unconformity-related	SIMS	9
	6	Mistamisk	Labrador Trough	Syn-metamorphic	SIMS	9
	2	Moore lake	Athabasca Basin	Intrusive	SIMS	9
	2	Sue	Athabasca Basin	Unconformity-related	SIMS	9
Czech	6	Rozna	Bohemian Massif	Unconformity-related	ICP-MS	12
	3	Okrouhla Radoun	Bohemian Massif	Unconformity-related	ICP-MS	12
Denmark	12	Motzfeldt Centre	Greenland	Intrusive	LA-ICPMS	13
Finland	2	Luthi	Nummi-Pusula peninsula	Intrusive	SIMS	9
France	4	Bois Noirs	Massif Central	Vein	LA-ICPMS	9
	2	Commanderie	Armorican Massif	Vein	LA-ICPMS	9
	2	Ecarpière	Armorican Massif	Vein	LA-ICPMS	9
Japan	4	Tono	Central Japan	Sandstone	SIMS	14
	28	Unazuki	Southwest Japan	Unconformity-related	SHRIMP	15
Namibia	2	Rössing	Damaran Belt	Intrusive	SIMS	9
Russia	2	Kola	Kola Peninsula	Intrusive	SIMS	9
	4	Streltsovkoye	Stretsovska Caldeira	Volcanic-related	SIMS	9
Zambia	6	Kawanga	Katanga System	Syn-metamorphic	SIMS	9

the multivariate data and implement REE patterns recognition. But PCA is an unsupervised multidimensional statistics technique and doesn't take sample classification into account. On the contrary, PLS-DA, which is a supervised technique and takes classification into account, is adopted to quantificationally characterize the groups, visualize the groups and discriminate how distinct the groups are when samples belong to different groups (i.e. origins), which provides a more meaningful interpretation on the point of nuclear forensic.

Sometimes, in the scatter of PLS-DA, there are extensive overlaps on the boundaries of near groups when outlier groups exist, which disturb the distinction of different groups. To solve the overlaps, iterative multidimensional statistics technique, a repetitive procedure of removing outlier groups and recalculating the model, has been introduced to attribution of UOCs and achieved satisfied results.[3,11] Moreover, another solution to the overlaps is application of advanced multidimensional statistics, such as orthogonal partial least squares discriminant analysis (OPLS-DA). Equipping with an additional treatment of multivariate data, OPLS-DA is an evolution of PLS-DA that has ability to separate between-group variances (e.g., samples are from different countries) from within-group variances (e.g., samples from a certain country have different types), which has much higher efficiency of the distinction of different groups.

In the normal case, before identifying the origin of an unknown UOC, a database should be established by gathering UOC samples from different origins. However, gathering UOC samples from all over the world is rarely practical and neither is referencing UOC data. To overcome the shortage of UOC samples, the aim of this study is to develop a novel approach to trace the origins of unknown UOCs using REE data of uranium ores instead of UOC samples.

The first part of this research is constructing a comparison database. Considering that gathering UOC samples or referencing UOC data is impossible for us, the feasible way to constructing the comparison database is referencing the REE data of uranium ores from published articles. The geographical and geological REE signatures of uranium ores are also applied to UOCs since REE pattern is invariable in milling processes. The second part is two-step pretreatment of the uranium ore data. The pretreatment consists of dimensionless treatment and logarithmic transformation, which can reduce the impact of big variances (e.g. REE levels in different uranium ores or different REE concentrations in a certain ore) to REE pattern recognition and reveal the fractionation of LREE and HREE more clearly. The third part is selection of suitable multidimensional statistics techniques. The pattern recognition efficiencies of PCA, PLS-DA and OPLS-DA are contrasted to select a suitable multidimensional statistics technique that can make different group separate and sub-groups within a certain group gather. The last part is establishing an attribution strategy. Employed iterative OPLS-DA, uranium ores from each country or deposit can be totally distinguished in a particular order which creates a decision tree, and the geographical and geological REE signatures of each group can be quantificationally characterized. Finally according to the decision tree, an unknown UOC can be attributed to its country and deposit and the type of uranium ore can also be analyzed.

2. Samples

The comparison database consisted of 116 uranium ores from 10 countries, 22 deposits (Table 1). All of the uranium ore REE data were referenced from published articles[9,12-15]. These uranium ores were classified by the classification of uranium ore[16]. Twelve REE compositions (La, Ce, Pr, Nd, Sm, Eu, Tb, Dy, Ho, Er, Tm and Lu) of the uranium ores were referenced. Because of interferences, Gd and Yb could not be measured by SIMS in the paper of J. Mercadier, et al. and they were not included in the comparison database. The REE concentrations under detection limits were taken as the value of detection limit.

Representativeness of the comparison database was ensured by: (1) all uranium ores were from 10 countries that possessed about 60% of the world uranium resource;[16] (2) these uranium ores covered six major uranium deposit types that represented about 50% of the world uranium resource;[16] (3) there were at least two deposit origins among the uranium ores from a vast country; (4) the number of uranium ores from each deposit was not less than two.

3. Method

3.1. Two-step pretreatment of data

Big variances might be present in both REE levels in different uranium ores and different REE concentrations in a certain uranium ore. Sometimes these variances could impact on the efficiency of REE pattern recognition. So before adopting multidimensional statistics techniques to pattern recognition, two steps of pretreatments were needed to reduce the impacts of these two kinds of variances respectively.

The first step was dimensionless treatment, which was transforming the concentration of a certain REE (except Tb) to the ratio of it and the concentration of Tb. Dimensionless treatment made the REE patterns of different uranium ores more comparable with each other and those of unknown UOCs. What's more, the indication of the geographical and geological signatures implied in the REE patterns enhanced. The second step was logarithmic transformation. Specifically, each REE ratio was taken to the logarithm of it to base 10, which gave all of the REE ratios relatively equal weights in pattern recognition. In addition, the reason why Tb was chosen as the divisor was: after logarithmic transformation, the positive and negative values of LREE or HREE ratio represented the concentration and depletion of LREE or HREE respectively, which could indicate the fractionation of LREE and HREE.

3.2. Selection of multidimensional statistics techniques

As mentioned earlier, a representative comparison database required that samples from some countries should have different deposits or types. However, there might be variances in these sub-groups, which compelled the group disperse in the scatter, caused overlapping with other groups and made it difficult to attribute. So except pretreatment of data, pattern recognition required a suitable multidimensional statistics technique which made different groups separate and sub-groups within a certain group gather. Obviously, PLS-DA and OPLS-DA were supervised techniques, and met the purpose of attribution better than PCA. OPLS-DA, an evolution of PLS-DA, could reduce the impact of within-group variances to pattern recognition on the point of algorithm, and was more suitable to attribution. In this study, the pattern recognition efficiencies of PCA, PLS-DA and OPLS-DA were compared with the help of multidimensional statistic software SIMCA 13.0 to select a suitable multidimensional statistics technique.

3.3. Establishment of an attribution strategy

There might be extensive overlaps on the boundaries of near

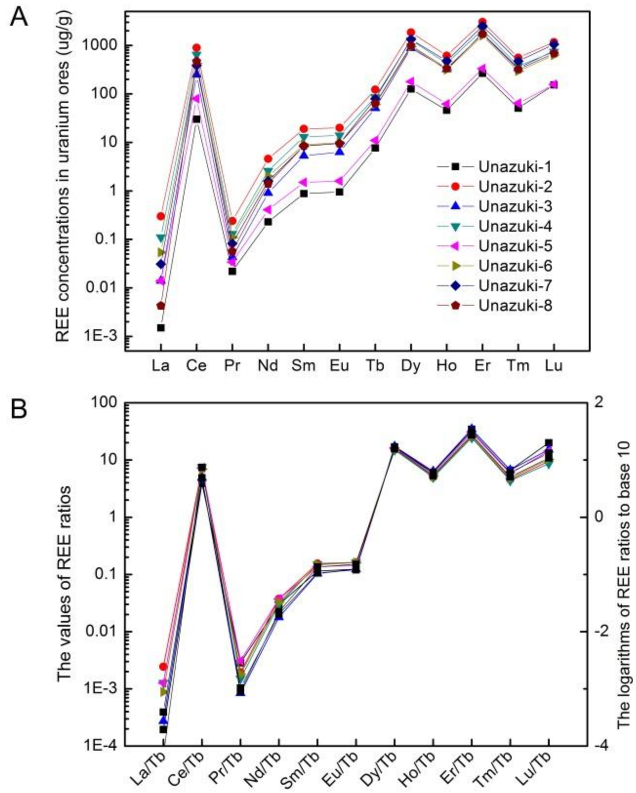


Fig.1 Two-step pretreatment of the REE data of the uranium ores from Unazuki, Japan

groups when an outlier group existed, because the variance between the outlier group and another group was so big that the other between-group variances could be ignored and those near groups mixed up in the scatter. The situation wouldn't improve even though an advanced multidimensional statistics technique had been adopted, and there was a demand to introduce iteration into the study and establish an attribution strategy.

For instance, an (or a few) outlier group(s) could be distinguished and the others mixed up in the scatter of the first iteration of OPLS-DA. The outlier group(s) would be removed and the rest groups (and the removed groups, if some of the removed groups mixed up) would be reanalyzed by another iteration of OPLS-DA. After several iterations, all groups were completely distinguished as far as possible, and the discrimination order of them formed a decision tree, i.e. attribution strategy.

The route of the strategized attribution of an unknown UOC should keep to the decision tree. Firstly the unknown UOC was added to the comparison database, and then the comparison database was analyzed by the first OPLS-DA. If the unknown UOC belonged to any of the outlier groups, the country of this UOC would be identified; if otherwise, the next iteration proceeded. The repetitive procedure would not stop until the unknown UOC was attributed to its country.

4. Results and discussion

4.1. Two-step pretreatment of data

The REE patterns of a small part of uranium ores from Unazuki, Japan before and after two-step pretreatment were shown in Fig

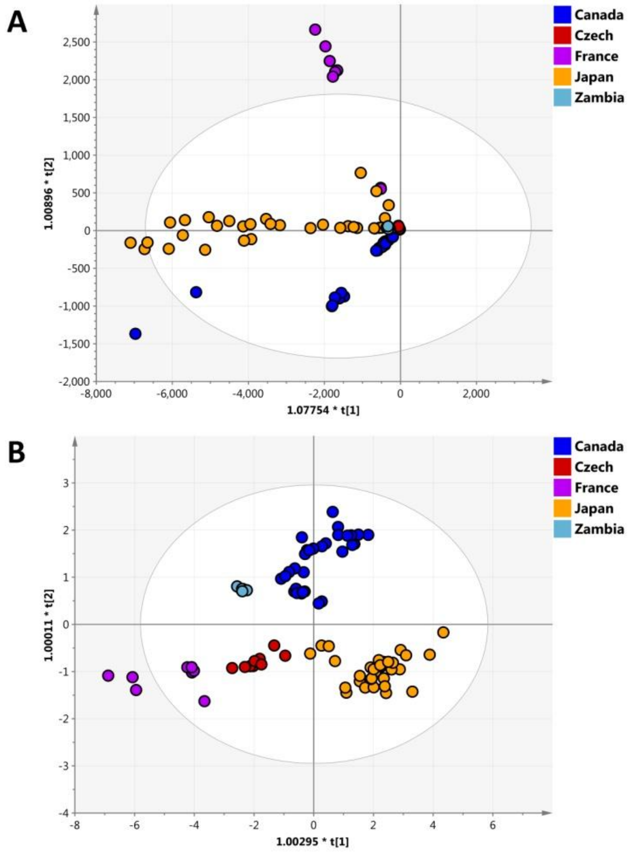


Fig.2 Comparison of the OPLS-DA scatters before and after two-step pretreatment of data

1A and B, respectively. Before pretreatment, there was a nearly hundredfold variance between the sample with the highest and lowest REE concentrations, which made the REE patterns of these samples incomparable. After dimensionless treatment, the REE patterns basically coincided, which showing that these samples came from a same geographical or geological origin. However, there was another problem that the highest REE ratio was about five orders higher than the lowest one. Thus the REE ratios were taken to the logarithm of them to base 10, and all of them were at the same order and possessed relatively equal weights when quantificationally characterizing the geographical and geological signatures by OPLS-DA. In addition, the negative values of LREE ratios and positive values of HREE ratios indicated the depletion of LREE and concentration of HREE.

To verify the necessity of two-step pretreatment, samples from some countries (Canada, Czech, France, Japan and Zambia) were analyzed by OPLS-DA before and after two-step pretreatment, which were shown in Fig 2A and B, respectively. After two-step pretreatment, the within-group dispersions and between-group overlaps caused by the big variances of REE levels in different uranium ores and different REE concentrations in a certain uranium ore were weakened to a certain degree, as a result the common feature of REE patterns in the uranium ores from a certain origin enhanced with the help of the two-step pretreatment.

4.2. Selection of multidimensional statistics techniques

To select a suitable multidimensional statistics technique, the scatters of PCA, PLS-DA and OPLS-DA analyzing the samples

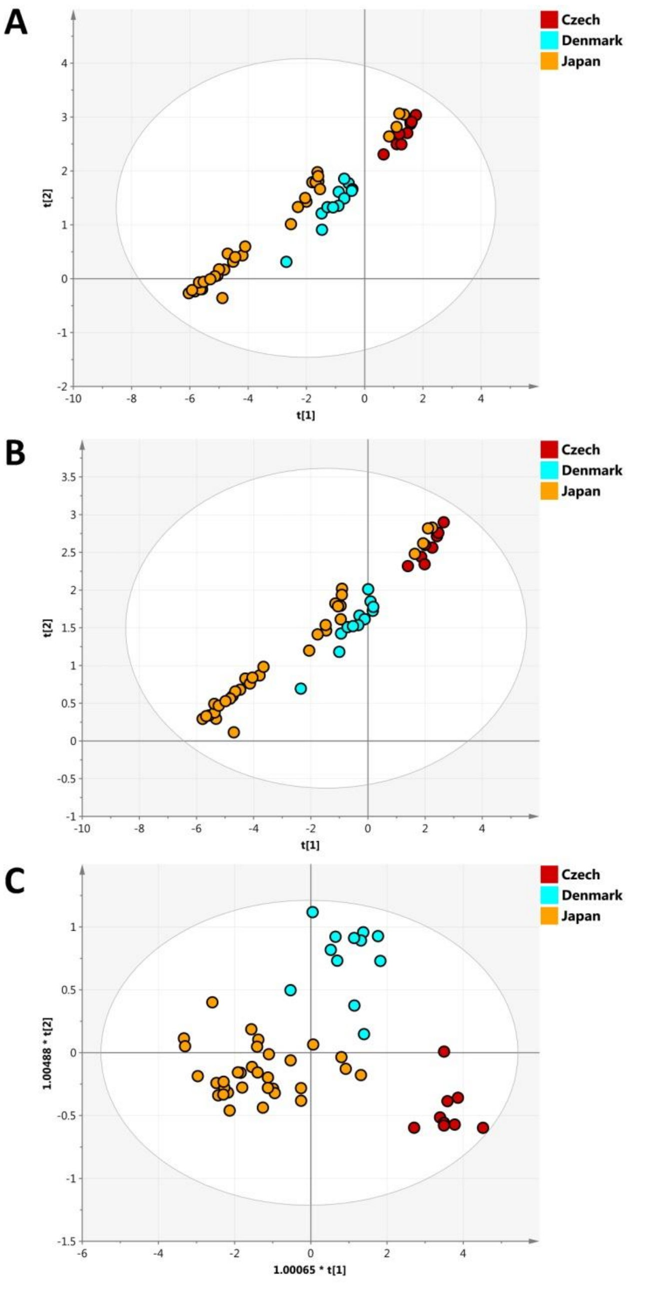


Fig.3 The scatters of PCA, PLS-DA and OPLS-DA analyzing the samples from four countries

from three countries (Czech Denmark and Japan) were contrasted and shown in Fig 3A, B and C, respectively. In the scatters of PCA and PLS-DA, samples from Japan were separated into three sub-groups. But in the scatter of OPLS-DA, these sub-groups gathered to form a group that could be distinguished with the other countries. Thus it could be verified that OPLS-DA obviously reduced the impact of within-group variances to between-group variances in pattern recognition on the point of algorithm, and OPLS-DA was a more promising multidimensional statistics technique than PLS-DA and PCA in the study of nuclear forensics.

4.3. Establishment of an attribution strategy

As mentioned in 3.3., iterative OPLS-DA was introduced to

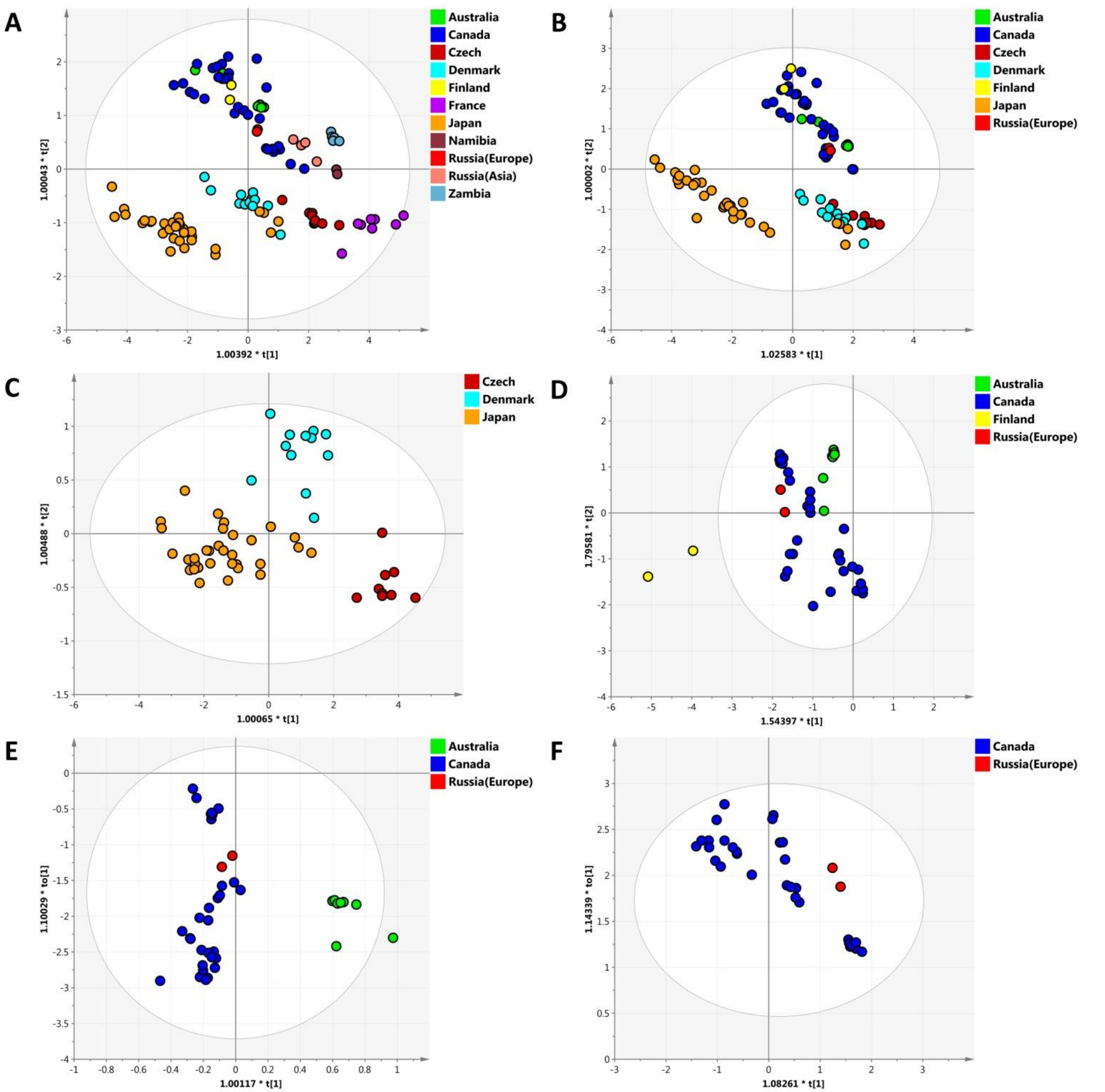


Fig.4 The scatters of six OPLS-DA iterations analyzing samples from ten countries

establish an attribution strategy. After two-stage pretreatment of the REE data in comparison database, the samples from ten different countries were analyzed by six iterations of OPLS-DA which were shown in Fig 4A, B, C, D, E and F. The scatter of the first iteration showed that the groups of France, Namibia, Russia (Asia) and Zambia were separated from the other countries. After taking these four groups away, the second OPLS-DA proceeded, and the left formed two mixed-groups: (1) Czech, Denmark and Japan; (2) Australia, Canada, Finland and Russia(Europe). In the third iteration, samples from Czech, Denmark and Japan were totally discriminated when analyzing the first mixed-group. Then the second mixed-group was analyzed in the fourth iteration, and the group of Finland was distinguished. After removing the samples from Finland, the group of Canada were also differentiated from Australia and Russia (Europe) in the fifth and sixth iteration, respectively. As a result, it showed that each group could be totally distinguished within several iterations and the common feature of REE pattern in each group could be quantificationally characterized in its iteration.

Thus a decision tree, i.e. attribution strategy, could be constructed with the discrimination orders of all groups, as was shown in Fig 5. If an unknown UOC needed to be identified to its country, the route of attribution should keep to the decision tree. The validity of the attribution strategy had been demonstrated by

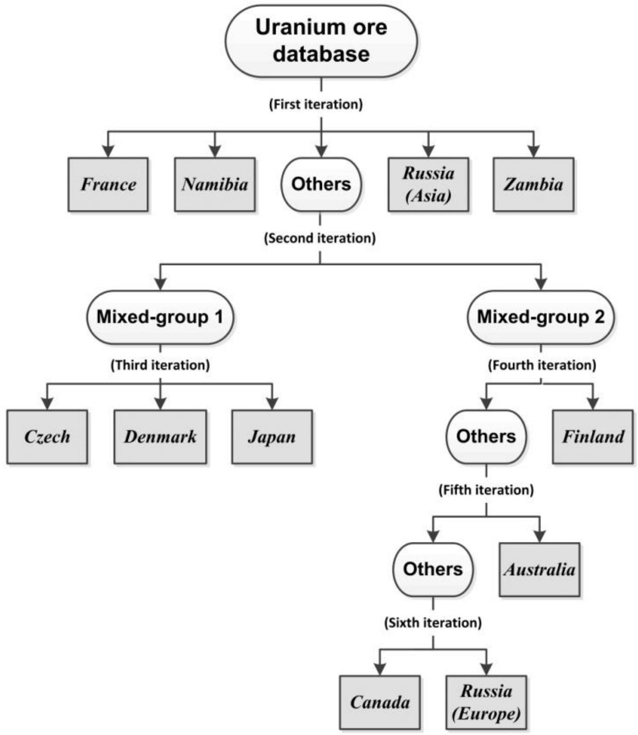


Fig.5 The decision tree constructed with the discrimination orders of all groups

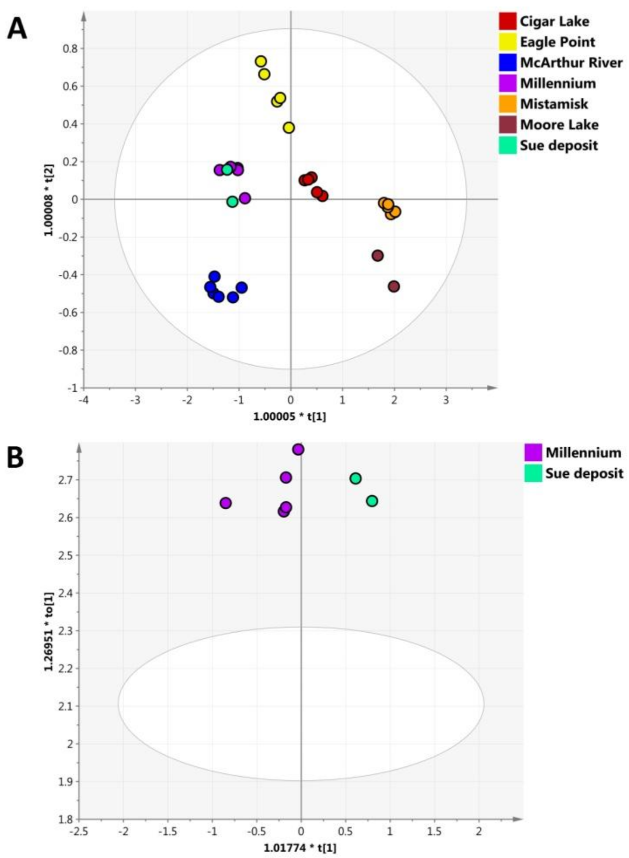


Fig.6 The scatters of two OPLS-DA iterations analyzing samples from Canada

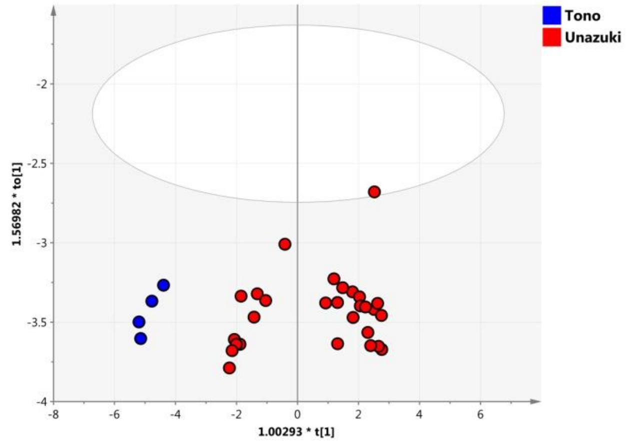


Fig.7 The scatters of OPLS-DA analyzing samples from Japan

Table.2 The coefficient of REE ratios in the first principal component in Fig.7

Atomic number	REE Ratio	Coefficient
57	La/Tb	-0.619
58	Ce/Tb	-0.133
59	Pr/Tb	-0.510
60	Nd/Tb	-0.444
62	Sm/Tb	-0.254
63	Eu/Tb	-0.171
66	Dy/Tb	0.052
67	Ho/Tb	0.074
68	Er/Tb	0.106
69	Tm/Tb	0.112
71	Lu/Tb	0.142

attributing several 'unknown' UOCs to their country according to the decision tree and the results were encouraging.

4.4. Identifying the deposit and type

Once the unknown UOC was attributed to a vast country, it was possible to determine the deposit of uranium ore by analyzing the samples from its country. For instance, the samples from Canada were classified by their deposits and calculated by OPLS-DA in Fig 6. The scatter showed that each group was distinguished with the others, except samples from Sue and Millennium deposits. And a next iteration indicated that samples from Sue and Millennium deposits had different common features of REE pattern. Obviously, an unknown UOC from Canada could be attributed to its deposit by this mean with the help of OPLS-DA.

Except the deposit, another attempt was made to identify the uranium ore type. For example, there were two uranium deposit types in the samples from Japan, among which the sandstone deposit (Tono) belonged to low-temperature deposit and unconformity-related deposit (Unazuki) was high-temperature deposit.[9] The scatter in Fig 7 showed that the high-temperature deposit were on the right of the low-temperature one, and the coefficient of REE ratios in the first principal component indicated that the coefficient increased with the atomic number as was shown in Table 2. All of these meant that there might be different fractionations of LREE and HREE in the high-temperature deposits and low-temperature deposits.

Specifically, in the high-temperature deposit, HREEs were concentrated since: (1) uranium deposited with the reduction of U (+6) to U (+4) in the igneous and metamorphic processes;[8] (2) HREEs (+3) had similar ionic radii with U(+4) so that they could substitute U(+4) in mineralization process[9]. It was just the opposite that LREEs were enriched in the low-temperature deposits because: (1) uranium migrated with underwater at low temperature and oxidized condition as soluble U (+6) complex and precipitated as insoluble U (+4); [8] (2) HREEs (+3) and anions (such as carbonate ion, fluorinion, etc) could form stable complexions[17] so that HREEs might migrate with water and LREEs might precipitate with U (+4).

The results proved that OPLS-DA was a powerful tool to reveal the fractionations of LREE and HREE, which contributed to identify the type of uranium deposit. As a consequence, it was a feasible way to attribute the unknown UOC to its deposit or type of uranium ore by analyzing the common feature of REE pattern implied in uranium ores. The REE pattern of uranium ore was a strong geographical and geological signature in attribution of UOC.

5. Conclusions

The study demonstrates that the geographical and geological signatures implied in the REE pattern of uranium ore can be used to attribute an unknown UOC to its origin (country, deposit or uranium deposit type) by iterative OPLS-DA. The results indicate that the within-group dispersions and between-group overlaps can be weakened by two-step pretreatment of data and application of OPLS-DA, and the efficiency of discrimination of different groups is greatly raised. A decision tree, or attribution strategy, constructed by iterative OPLS-DA can distinguish samples from ten different countries. It is feasible to attribute an 'unknown' UOC to its country according to the decision tree and the deposits or deposit types might also be identified by analyzing samples from its country. Overall, unknown UOCs can be attributed to their origins through a novel approach using REE data of uranium ores instead of UOC samples and iterative OPLS-DA is a promising approach for the studies of nuclear forensics.

Notes and references

[a] *Department of Radiochemistry, China Institute of Atomic Energy, Box 8, P.O. BOX 275, Beijing, China. Fax: +86-10-69357936; Tel: +86-10-69358448; E-mail: zhaoyg@ciae.ac.cn*

[b] *State Nuclear Power Research Institute, Beijing, China. Fax: +86-10-56681000, Tel: +86-10-56681825, E-mail: linmu@snptc.com.cn*

* *Corresponding author. This work was finished in China Institute of Atomic Energy.*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

[1] IAEA, *Nuclear forensics support, IAEA nuclear security series No. 2*, Vienna, 2006.
[2] V. Badaut, M. Wallenius and K. Mayer, *J. Radioanal. Nucl. Ch.*, 2009, 280, 57-61.
[3] E. Keegan, M. Wallenius, K. Mayer, Z. Varga and G. Rasmussen, *Appl. Geochem.*, 2012, 27, 1600-1609.
[4] E. Keegan, S. Richter, I. Kelly, H. Wong, P. Gadd, H. Kuehn and A. Alonso-Munoz, *Appl. Geochem.*, 2008, 23, 765-777.
[5] Z. Varga, M. Wallenius, K. Mayer and M. Meppen, *Radiochim. Acta.*, 2011, 1, 1-4.
[6] Z. Varga, M. Wallenius, K. Mayer, E. Keegan and S. Millet, *Anal. Chem.*, 2009, 81, 8327-8334.
[7] S. H. Han, Z. Varga, J. Krajkó, M. Wallenius, K. Song and K. Mayer, *J. Anal. At. Spectrom.*, 2013, 28, 1919-1925.
[8] G. A. Brennecka, L. E. Borg, I. D. Hutcheon, M. A. Sharp and A. D. Anbar, *Earth. Planet. Sc. Lett.*, 2010, 291, 228-233.
[9] J. Mercadier, M. Cuney, P. Lach, M. C. Boiron, J. Bonhoure, A. Richard, M. Leisen and P. Kister, *Terra Nova*, 2011, 23, 264-269.
[10] Z. Varga, M. Wallenius and K. Mayer, *Radiochim. Acta.*, 2010, 98, 771-778.
[11] M. Robel and M. J. Kristo, 2nd Workshop on Nonproliferation Nuclear Forensics, Beijing, China, 2010.
[12] M. Rene, *Terra Nova*, 2008, 20, 52-58.
[13] J. A. McCreath, A. A. Finch, S. L. Simonsen, C. H. Donaldson and A. Armour-Brown, *Contrib. Mineral. Petr.*, 2012, 163, 967-982.
[14] Y. Takahashi, H. Yoshida, N. Sato, K. Hama, Y. Yusa and H. Shimizu, *Chem. Geol.*, 2002, 184, 311-335.
[15] K. Horie, Y. Tsutsumi, M. Cho, Y. Morishita and H. Hidaka, Phys. Chem. Earth., 2010, 35, 284-291.
[16] IAEA. *World Distribution of Uranium Deposits (UDEPO) with Uranium Deposit Classification*, *IAEATECHDOC-1629*, Vienna, 2009.
[17] S. Göb, J. E. Gühring, M. Bau and G. Markl, Am. Mineral., 2013, 98, 530-548.