

Titolo
Aspetti metodologici per la sicurezza delle sorgenti radioattive
GALAXY SERPENT EXERCISE v3
Descrittori

Tipologia del documento: RAPPORTO TECNICO

Collocazione contrattuale: Accordo di programma ENEA-MSE su sicurezza nucleare e reattori di IV generazione.

Argomenti trattati: metodologie, security, nuclear forensic, sorgenti radioattive.

Sommario

Il Nucleare Forense è una delle colonne portanti nell'architettura della "Nuclear Security", come risposta ad atti criminali e traffici illeciti che coinvolgano materiali nucleari o altri materiali radioattivi. Nell'ambito di tale tematica e coerentemente al tema A 2.2, ENEA ha partecipato ad un esercizio internazionale indetto dalla comunità forense con lo scopo di aiutare lo sviluppo di competenze nazionali. Il "GALAXY SERPENT" è un insieme di esercizi internazionali virtuali web-based, focalizzati sullo sviluppo di librerie forensi nucleari nazionali (NNFLs). Il fine è incrementare la consapevolezza nazionale riguardo le sfide tecniche associate con lo sviluppo di una NNFL e come essa possa costituire un valido supporto per le investigazioni di crimini e/o attività illecite che coinvolgano materiali nucleari o radioattivi. In questo particolare caso, il "GALAXY SERPENT EXERCISE v.3", è stato dedicato ad un particolare materiale radioattivo, il concentrato d'Uranio. L'esercizio in questione è stato strutturato in due fasi. La prima fase (Phase 1) è consistita nello sviluppo di una componente della NNFL, dedicata a matrici geologiche contenenti concentrato d'Uranio, a partire dal database fornito. La seconda fase (Phase 2), il cui obiettivo è stato la verifica dell'efficacia e l'utilizzo della libreria sviluppata in un contesto semplificato di analisi forense simulata, è stata incentrata sullo scenario virtuale dell'ipotetica intercettazione di campioni geologici radioattivi, con lo scopo di verificare l'eventuale corrispondenza con i materiali radioattivi contenuti nella NNFL.

Note

Autori: Giuseppe Ottaviano, Franca Padoani, Antonietta Rizzo, Federico Rocchi, Chiara Telloli.


Copia n.

In carico a:

2			NOME			
			FIRMA			
1			NOME			
			FIRMA			
0	EMISSIONE	23/11/18	NOME	Giuseppe Ottaviano	FRANCA PADOANI	FEDERICO ROCCHI
			FIRMA			
REV.	DESCRIZIONE	DATA		REDAZIONE	CONVALIDA	APPROVAZIONE

INDICE

1	INTRODUZIONE.....	3
2	STRUTTURA DELL'ESERCIZIO E DESCRIZIONE DELLE FASI	3
2.1	FASE 1: SVILUPPO DELLA NNFL.....	4
2.2	FASE 2: SCENARIO - INTERCETTAZIONE DI CAMPIONI RADIOATTIVI, VERIFICA DELLA CONSISTENZA RECIPROCA E VERIFICA DELLA CONSISTENZA CON LA NNFL	4
3	"MODUS OPERANDI" DEL GRUPPO DI LAVORO ENEA.	4
3.1	FASE 1 - SVILUPPO DELLA NNFL	6
3.2	FASE 1 - RISULTATI.....	10
3.3	FASE 2 - SCENARIO: INTERCETTAZIONE DI CAMPIONI RADIOATTIVI, VERIFICA DELLA CONSISTENZA RECIPROCA E VERIFICA DELLA CONSISTENZA CON LA NNFL	10
3.4	FASE 2 - RISULTATI.....	12
4	SOLUZIONE DELL'ESERCIZIO.....	12
4.1	FASE1.....	13
4.2	FASE 2	13
5	CONFRONTO TRA LE SOLUZIONI	13
5.1	FASE 1.	13
5.2	FASE 2.	13
6	CONCLUSIONI	14
7	ALLEGATO A: DETTAGLI STATISTICI FASE 1.	15
8	ALLEGATO B: SUMMARY REPORT ON GSV3 – PHASE 1.....	17
9	ALLEGATO C: GALAXY SERPENT V3 – SUMMARY REPORT – PHASE 2.....	36
10	BIBLIOGRAFIA.....	49

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	3	49

1 INTRODUZIONE

Il “GALAXY SERPENT” è un insieme di esercizi internazionali virtuali web-based, focalizzati sullo sviluppo di librerie forensi nucleari nazionali (NNFL’s).

Il primo esercizio “GALAXY SERPENT exercise v.1” ha avuto luogo nel 2013 ed è consistito nella verifica della consistenza di un ipotetico combustibile nucleare esausto confiscato, con i dati nella NNFL sviluppata da ciascun partecipante.

Il secondo esercizio, il “GALAXY SERPENT exercise v.2”, è stato dedicato alle sorgenti radioattive sigillate ed è stato organizzato e gestito con la creazione di un portale dedicato, il “Central Desktop”, attraverso il quale sono state fornite tutte le informazioni necessarie. L’esercizio v.2 ha avuto inizio il 27 luglio 2015 con l’inserimento, sul “Central Desktop”, dei documenti di riferimento e dei dati (injects) necessari e sufficienti per affrontare l’esercizio stesso e ha avuto termine con la pubblicazione della soluzione in data 18 Febbraio 2016 [ADPFISS-LP1-082] (documento ANL-DHS-2016-007).


Il terzo esercizio, il “GALAXY SERPENT exercise v.3” è stato dedicato a materiali radioattivi di origine geologica contenenti Uranio (Uranium Ore Concentrate, UOC) ed è stato organizzato e gestito con la creazione di uno spazio dedicato sul portale Slack, attraverso il quale sono state fornite tutte le informazioni necessarie. L’esercizio v.3 ha avuto inizio in data 1 giugno 2017 con l’inserimento, sul portale Slack, dei documenti di riferimento e dei dati necessari e sufficienti per affrontare l’esercizio stesso e ha avuto termine con la pubblicazione, in data 13 giugno 2018, del “Galaxy-Serpent Phase 2 Summary Report”. La partecipazione di ENEA a questo esercizio è importante poiché s’inserisce all’interno delle attività già da tempo programmate e realizzate nell’ambito della sicurezza e della mitigazione del rischio chimico, biologico, radiologico, nucleare ed informatico (CBRNe). L’esercizio, pertanto, ha avuto il duplice scopo di testare da un lato il livello di competenze e il tipo di strumenti necessari a una risposta in ambito CBRN, e dall’altro di predisporre strumenti e procedure che possono vedere interessati anche altri attori a livello nazionale, in particolare nella risposta al rischio radiologico. Gli autori del rapporto, hanno portato a termine l’esercizio, contribuendo alla sua risoluzione grazie alle diverse competenze in fisica, chimica e ingegneria.

2 STRUTTURA DELL’ESERCIZIO E DESCRIZIONE DELLE FASI

L’esercizio v.3 è stato strutturato in due fasi.

La Fase 1 è stata caratterizzata da uno scenario virtuale incentrato sullo sviluppo di una NNFL dedicata alle sorgenti radioattive di origine geologica (Concentrato d’Uranio, UOC), a partire dal database fornito. Gli autori dell’esercizio hanno diffuso, attraverso il portale “Slack”, un file con la descrizione del primo scenario e un file con gli 821 campioni atti a costituire la NNFL. La Fase 2 è stata concepita per verificare l’efficacia e l’utilizzo della libreria sviluppata in un contesto semplificato di analisi forense simulata. Essa è consistita in uno scenario virtuale incentrato sulla intercettazione di alcuni campioni di materiale geologico radioattivo e ha avuto, quale obiettivo, la verifica dell’eventuale corrispondenza tra i campioni radioattivi intercettati e i campioni radioattivi costituenti la libreria creata nella prima fase. Gli autori dell’esercizio hanno inserito, sul portale “Slack”, un file con la descrizione del secondo scenario e un file con la specificazione dei tre campioni intercettati.

Le informazioni generali per la strutturazione della NNFL sono state estratte dai seguenti documenti:

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. di 4 49
--	--	------------------	----------------------	------------------------

- “IAEA NSS2 Nuclear Forensic Support Technical Guidance Reference Manual”;
- “Proposed Framework for National Nuclear Forensic Libraries and International Directories”, J. F. Wacker (Pacific Northwest National Laboratory), U.S. Department of Energy, Michael Curry (U. S. Department of State) – 08 June 2011;
- “IAEA NSSXX DEVELOPMENT OF A NATIONAL NUCLEAR FORENSICS LIBRARY”, DRAFT, February 2013.

2.1 Fase 1: sviluppo della NNFL

La Fase 1 ha descritto una attività di importazione di concentrato d’Uranio da uno stato fittizio, Pacifica. Il governo richiede l’organizzazione delle informazioni riguardanti i materiali importati, in una Libreria Forense Nucleare Nazionale (NNFL), al fine di supportare eventuali indagini investigative forensi.

2.2 Fase 2: scenario - intercettazione di campioni radioattivi, verifica della consistenza reciproca e verifica della consistenza con la NNFL

La Fase 2 ha descritto uno scenario d’intercettazione, in seguito ad anonimo suggerimento, di un veicolo che sta cercando di lasciare il paese con un carico illecito. Individuati a bordo tre contenitori sigillati e marcati unicamente con simboli di allerta-radioattività. Le ispezioni condotte rivelano la presenza di sorgenti geologiche in ciascun contenitore e le successive indagini isotopiche specificano la presenza di concentrato d’Uranio. Il governo richiede quanto segue:

1. Verificare la reciproca consistenza dei campioni intercettati.
2. Verificare la consistenza tra i campioni intercettati e i campioni costituenti la NNFL stessa.
3. Indicare le azioni da intraprendere dopo aver risposto ai precedenti quesiti.
4. Indicare il modo di condividere i risultati all’interno del governo stesso.

3 “MODUS OPERANDI” DEL GRUPPO DI LAVORO ENEA.

Il Gruppo di Lavoro ENEA che ha partecipato all’esercizio è stato scelto sulla base delle diverse competenze (fisica, chimica, ingegneria) e sull’esperienza in campo radiologico e di campionamento ambientale. I ricercatori partecipanti sono tutti coinvolti nelle attività di security, safety e mitigazione del rischio CBRNe, della “Divisione Sicurezza e Sostenibilità del Nucleare” del “Dipartimento Fusione e Tecnologie per la Sicurezza Nucleare”.

La scelta operativa per le attività dell’esercizio ha previsto che ciascun ricercatore, singolarmente, studiasse il materiale proposto nelle tre singole fasi e, per ogni fase, è stata organizzato un incontro collegiale per lo scambio di commenti ed osservazioni.

Gli obiettivi realizzativi di tali incontri collegiali per ciascuna fase sono stati i seguenti:

- Fase 1:
 - Progettazione di un database dedicato alla raccolta dei dati relativi ai materiali radioattivi forniti dall’esercizio
 - Scelta della piattaforma informatica da utilizzare per il database (sono state proposti Excel, Access, FileMaker). La piattaforma prescelta è stata Excel per la buona conoscenza di tale strumento da parte di tutti i componenti del gruppo

e per la facilità di scambio dei relativi documenti prodotti. In questa fase la scelta dello strumento, con cui progettare l'archivio, è stata guidata in maniera predominante dalla necessità di uno scambio d'informazioni pratico e funzionale e dalla comune conoscenza della struttura del database. L'esigenza di una piattaforma comune, ben nota a tutti, è fondamentale nel caso reale di una situazione di emergenza in cui, verosimilmente, si troverebbero a lavorare insieme esperti con diverse competenze e livelli di istruzione informatica.

- Scelta dei parametri interrogabili da inserire nel database (cioè filtrabili) e scelta delle informazioni che non necessitavano interrogazione (es. dati chimico fisici, tempi di emivita dei radioisotopi, costanti fisiche, ...) e che pertanto sono state inserite nel database come singoli fogli elettronici non interconnessi con il database interrogabile.

➤ Fase 2:

- Discussione e scelta della metodologia più idonea per interrogare la NNFL e quindi verifica della consistenza del materiale intercettato con i campioni costituenti la libreria stessa. Questa è stata la fase più critica e complessa, in quanto si sono dovute integrare le diverse competenze, sensibilità e esperienze dei ricercatori coinvolti, cercando di focalizzarsi sulla strategia più obiettiva. È stato interessante notare come gli approcci utilizzati siano effettivamente legati, non solo alle competenze acquisite, ma anche alle esperienze lavorative degli esperti coinvolti, per cui è importante che questi tipi di esercizi prevedano anche delle attività sperimentali dedicate. La II fase è stata caratterizzata da frequenti confronti per verificare la correttezza del metodo applicato

Durante la Fase 2, la strategia scelta per il completamento delle determinazioni analitiche è consistita nell'affidare ad un singolo ricercatore l'esecuzione delle stesse e poi confrontarsi periodicamente per verificare i risultati. Tale scelta ha permesso, da un lato, di avere un esperto completamente dedicato al trattamento analitico dei dati garantendo quindi omogeneità e congruenza delle valutazioni, dall'altro di minimizzare gli errori sistematici e casuali attraverso la verifica indipendente degli altri esperti.

Nelle due fasi menzionate, la valutazione e la scelta delle strategie di interrogazione del database, invece, è stata guidata dal confronto e l'integrazione di tutte le osservazioni degli esperti, senza che ci fosse un esperto "guida" scelto a priori. Tale strategia ha permesso di verificare l'importanza dell'integrazione delle competenze e di assicurare la maggiore obiettività possibile nella risposta all'esercizio.

La NNFL può essere assimilata a una matrice $M \times N$, in cui le M righe identificano gli M campioni attraverso un codice alfanumerico identificativo e la classe di affinità mentre le N colonne sono costituite dagli elementi che costituiscono i campioni e da informazioni di localizzazione; in tal modo ogni campione costituisce un punto nello spazio N -dimensionale degli elementi (Figura 1):

NNFL (matrice MxN)

Campioni (affinità+codice)	E ₁	E ₂	E ₃	E _N
C ₁													
C ₂													
C ₃													
...													
...													
...													
...													
C _M													

Tabella 1: schema di principio della matrice NNFL.
3.1 Fase 1 - Sviluppo della NNFL

Il Gruppo di Lavoro ENEA ha dunque acquisito le informazioni, diffuse dagli autori dell'esercizio internazionale attraverso il portale dedicato "Slack", per la costituzione (Fase 1) della "Libreria Forense Nucleare Nazionale" (NNFL) e per affrontare le successive fasi dell'esercizio.

La matrice della NNFL è stata sviluppata creando un file Microsoft Excel, ordinando i campioni secondo le quattro classi di affinità previste ossia i quattro tipi di basalto (IAB, MORB, OIB, ZCRFB) da cui è stato estratto il Concentrato d'Uranio e ordinando gli elementi secondo il numero atomico. Ciascuna classe di affinità comprendeva alcune centinaia di campioni e per ciascun campione sono state riportate le seguenti informazioni:

- codice alfanumerico identificativo
- latitudine e longitudine
- composizione elementare (nessuna specificazione isotopica)

La matrice della NNFL è dunque costituita da 821 righe e 49 colonne e cioè 45 colonne delle composizioni elementali più quattro colonne con le informazioni di localizzazione, classi di affinità e codici alfanumerici (Figura 2). Per le successive analisi, si è ritenuto di poter trascurare le colonne riportanti le informazioni di localizzazione e pertanto la matrice della NNFL è risultata costituita da 821 righe e 47 colonne. Presenti, dunque, 821 campioni e 45 elementi con concentrazioni espresse in ppm, tranne che per l'Uranio i cui valori sono stati espressi in wt%. Per uniformare le unità di misura si è proceduto a convertire la concentrazione dell'Uranio in ppm.

A	B	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
1	AFFINITY	SAMPLE_ID	U	Ti	Al	Fe	Ca	Mg	Mn	K	Na	P	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
2	IAB	58685	788000	0,11	11,94	9,2	11,44	10,39	0,18	1,52	2,81	0,4	39,3	44,2	7,39	22,3	3,69	0,13	3,12	0,68	5,13	1,58	8,21	2,01	20,4	3,5
3	IAB	20972-FIL_93	809000	0,42	19,4	8,5	10,7	5,5	0,16	1,3	2,3	0,21	46,5	86,4	8,88	24,3	5,8	0,16	6,47	1,61	12,7	3,15	12,3	2,02	13,4	1,63
4	IAB	29358-P.1147.3	796000	0,30	19,6	8,17	12,22	7,4	0,14	0,38	2,04	0,08	27,4	96,7	12,6	4,2	11,4	0,26	11,4	2,32	14,5	2,87	9,11	1,3	8,82	1,27
5	IAB	12990-AB953-1	815000	0,41	18,39		10,25	4,83	0,18	1,15	2,47	0,19	24,5	69,2	11,8	39,9	11,2	0,24	11,9	2,28	14,3	2,83	8,8	1,18	8,21	1,17
6	IAB	14662-WA48	820000	0,27	16,66		8,89	9,12	0,12	0,41	3,84	0,04	45,7	98,5	8,66	23,8	5,81	0,17	6,59	1,66	3,3	3,3	12,8	2,17	15,1	1,93
7	IAB	48990	793000	0,42	17,59		12,07	6,2	0,19	0,56	2,24	0,1	54,4	122	10,9	31,5	5,02	0,14	4,55	0,92	7,25	2,37	13,4	3,45	37,3	6,45
8	IAB	61883	795600	0,50	19,49		11,22	5,5	0,15	0,24	2,19	0,12	88,9	150	17,8	58,6	11,6	0,55	10	2,17	14	3,18	11,1	1,66	11,6	1,51
9	IAB	170120	815000	0,50	19,8		10,02	5,54	0,19	0,68	2,9	0,17	131	230	27,4	91,9	18,8	1,05	18,5	3,93	26,3	5,78	21,3	3,06	19,9	2,62
10	IAB	8399-24	769500	0,50	17,6	8,74	10,28	11,59	0,21	0,59	2,22	0,14	147	241	26	81,5	15,9	0,56	14,9	3,21	21	4,67	16	2,39	15,9	2,16
11	IAB	10199-1072	818400	0,19	16,79		12,16	6,79	0,18	0,16	1,19	0,02	28,6	80	12,7	42,2	11,1	0,22	10,1	2,09	12,1	2,35	6,89	0,95	6,79	0,97
12	IAB	90543	771000	0,50	19,2	9,99	7,54	6,01	0,2	0,89	3,44	0,08	103	143	18,6	63,5	12,4	0,67	13,2	2,57	18	4,06	14,6	2,13	14,2	1,83
13	IAB	43502	816300	0,43	19,63	9,13	10,01	5,56	0,16	0,21	2,12	0,08	204	300	30,8	86,9	14,8	0,46	12,9	2,6	15,4	3,34	11,9	1,59	10,4	1,4
14	IAB	137268	803900	0,50	18,44	10,88	10,05	6,07	0,19	0,22	2,11	0,06	243	271	33,9	96,3	16,9	0,53	16,1	3,02	19,4	4,23	14	1,87	11,9	1,58
15	IAB	51014	816200	0,26	16,65		9,93	3,63	0,2	0,4	2,27	0,156	104	168	20,7	68,5	15,1	0,28	14,9	3,01	20,4	4,26	14	1,89	12,9	1,72
16	IAB	142652	812700	0,50	17,45	7,86	10,42	7,8	0,17	0,61	2,89	0,11	295	368	55,6	183	32,8	1,9	32,8	5,54	35,3	7,56	25	3,42	22,1	2,83
17	IAB	8801-SST.6.3	813700	0,41	16,15		11,47	6,19	0,22	0,29	2,19	0,05	126	196	22,5	73,6	15,9	0,3	17,2	3,26	21,2	4,62	15,4	2,12	14	1,91
18	IAB	8197-SSM.5.4	786600	0,28	18,58		12,89	8,24	0,18	0,1	1,3	0,35	138	228	23,5	72,1	15,3	0,37	16,4	3,61	25,7	6,11	22,7	3,36	22,7	3,05
19	IAB	8790-SSC.35.1	803700	0,41	18,65		11,65	5,78	0,18	0,055	1,82	0,02	176	306	27,9	87,1	19,9	0,48	24,1	4,92	36	8,45	30,6	4,6	31,7	4,17
20	IAB	14182-SSS_10.1	803400	0,50	16,83	9,71	11,31	7,45	0,17	0,35	2	0,09	241	394	48,7	167	31,5	1,82	27,4	5,51	34,2	7,39	25	3,55	22,2	2,91
21	IAB	41026	798200	0,31	19,86		12,38	6,59	0,18	0,09	1,71	0,03	231	469	56,8	178	35,9	1,16	36	6,77	39,8	8,03	24,7	3,11	19,1	2,3
22	IAB	40947	799200	0,34	16,53		12,58	8,27	0,18	0,12	1,6	0,02														
23	IAB	41013	803300	0,40	19,04		11,73	6,08	0,2	0,12	1,67	0,03														
24	IAB	41014	803400	0,40	18,81		11,7	6,27	0,19	0,12	1,68	0,03														
25	IAB	41064	806200	0,44	17,58		11,11	6,12	0,17	0,17	2,14	0,07														
26	IAB	42382	800100	0,45	20,18		11,36	3,86	0,18	0,42	2,35	0,11														
27	IAB	43108	798700	0,44	16,31	9,96	11,74	6,43	0,15	0,6	2,35	0,16														
28	IAB	43498	807700	0,95	16,21	11,24	9,24	5,64	0,18	0,93	2,8	0,39														
29	IAB	43504	820100	0,62	17,47	11,58	10,41	4,21	0,19	0,39	1,71	0,09														
30	IAB	43568	830000	0,67	16,1	9,22	8,78	6,94	0,16	0,84	3,67	0,21														
31	IAB	43847	826200	0,52	15,55		10,37	6,22	0,22	0,15	1,47	0,05														
32	IAB	43849	811200	0,53	17,83		11,67	5,47	0,19	0,39	2,22	0,08														

Tabella 2: estratto dalla matrice NNFL (Fase 1).

Ognuno degli 821 campioni della libreria è rappresentabile da un punto in uno spazio a 45 dimensioni oppure, in altri termini, i 45 elementi definiscono uno spazio a 45 dimensioni popolato da 821 punti. Si è in presenza di un tipico problema multivariato definito da 45 variabili eventualmente correlate. Si è reso pertanto necessario ricorrere ad un approccio per gestire efficacemente un tale insieme di dati e cioè un approccio che consenta di trarre informazioni che spieghino e diano ragione dei dati stessi. L’approccio per gestire i dati contenuti nella matrice della NNFL è stato basato sull’utilizzo di una tecnica statistica avanzata, l’**Analisi delle Componenti Principali (ACP)**, eseguita per mezzo di MATLAB v. R2014a.

La ACP appartiene alle tecniche statistiche di analisi multivariata, ed è applicabile a insiemi di osservazioni iniziali eseguite al fine di disporre di un vasto ambito di informazioni, conseguenti a indagini condotte su campioni più o meno vasti. L’obiettivo dell’ACP consiste nell’individuare alcuni (pochi) fattori di fondo che spieghino e diano ragione dei dati stessi. Tali fattori di fondo, o componenti, rappresentano delle dimensioni “ideali” dotate di significato. La ACP realizza la possibilità di esprimere in forma sintetica insiemi di variabili tra loro correlate riducendo il numero delle variabili stesse. Si tratta, cioè, di ricombinare le misure individuando strutture lineari indipendenti che possano essere interpretate come indici non correlati.

I dettagli della ACP sono riportati in Appendice A.

Dalla matrice è stato eliminato l’Uranio per i seguenti motivi:

- Esso spiega circa il 95% della varianza conducendo così ad un risultato fortemente condizionato
- Esso presenta la minore variabilità rispetto a tutti gli altri elementi (vedi figura 3)
- Esso non è un “trace element”
- Esso è il principale componente dei campioni della NNFL e pertanto non è utile ai fini di una investigazione di tracciabilità

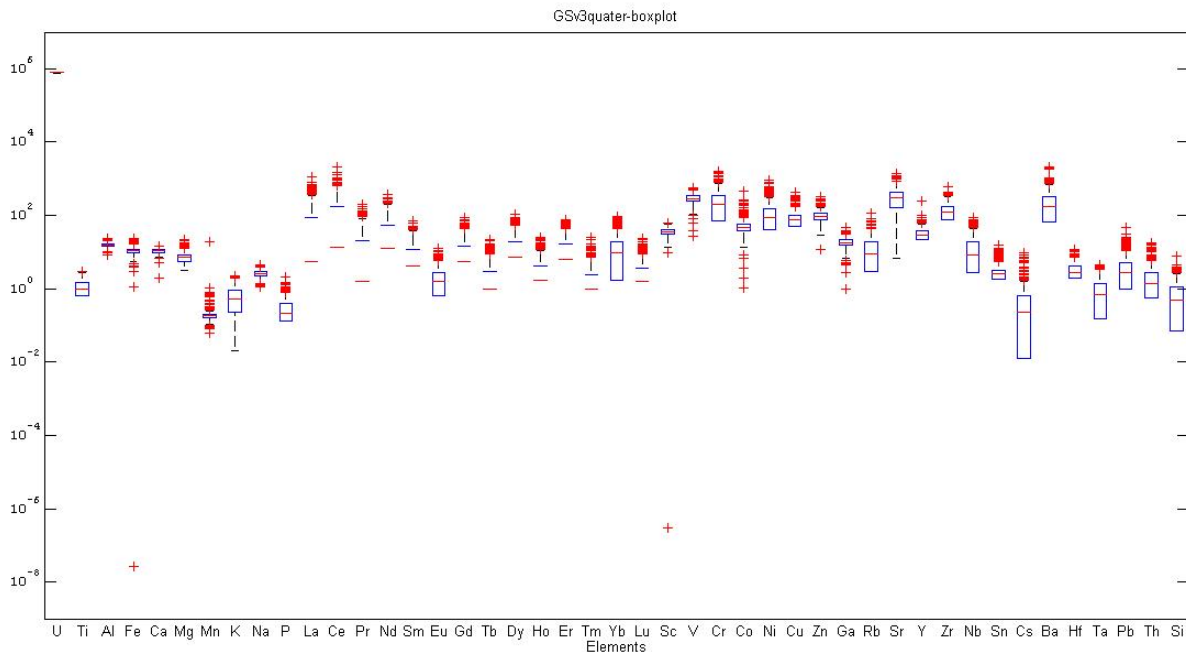


Figura 1: box-plot con la variabilità dei 45 elementi del database.

Di conseguenza, il numero delle colonne degli elementi è stato ridotto da 45 a 44 e così la ACP è stata portata a termine sulla matrice 821x44. Tale matrice presentava tuttavia dei dati mancanti e pertanto, prima della esecuzione della ACP, essi sono stati ricostruiti per mezzo di stime fisicamente neutre ottenute applicando l'algoritmo di MATLAB, Alternate Least Square (ALS). La successiva esecuzione della ACP ha mostrato che le prime cinque componenti spiegano circa il 95% della varianza, le prime tre componenti principali spiegano circa il 83% della varianza e le prime due componenti principali spiegano più dei due terzi della varianza ossia circa il 71% (vedi figure 4 e 5).

Variables - explained		
explained		
44x1 double		
	1	2
1	39.3256	
2	32.6248	
3	12.3935	
4	10.0768	
5	1.5667	
6	1.2122	
7	0.8796	
8	0.6104	
9	0.5245	
10	0.2495	

Tabella 3: valori della varianza spiegata dalle prime dieci componenti principali.

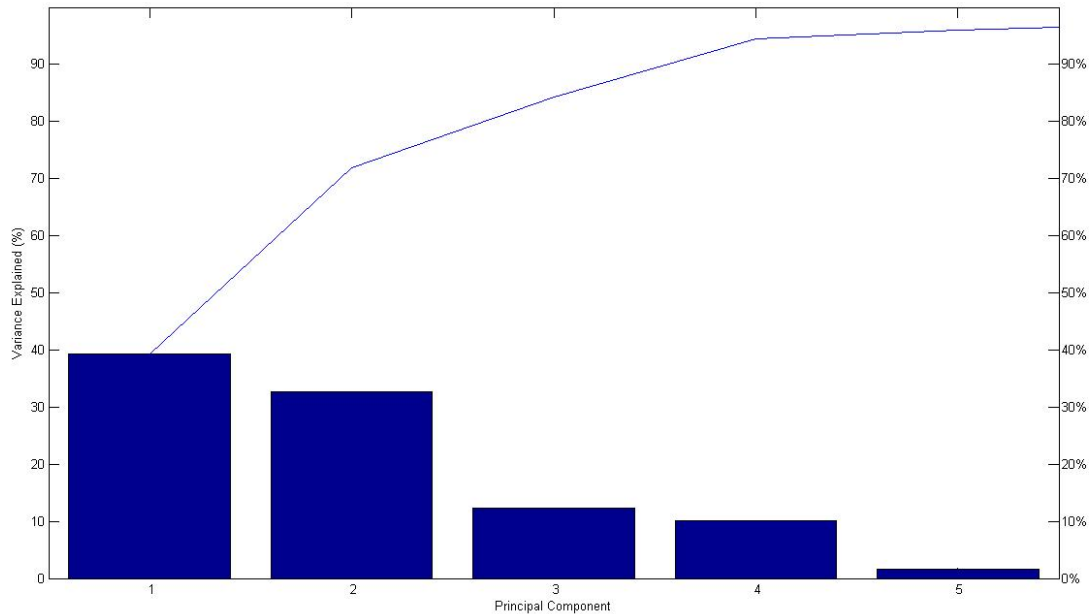


Figura 2: grafico a barre (screeplot) con la percentuale di varianza spiegata dalle prime cinque componenti principali.

E' stato pertanto ragionevole considerare le prime tre componenti principali in modo da ridurre le dimensioni dello spazio da 44 a 3 oppure anche le prime due componenti principali in modo da ridurre le dimensioni dello spazio da 44 a 2.

L'interpretazione dei risultati è stata resa possibile rappresentando in un unico grafico "bi-plot", gli 821 campioni e le 44 variabili. Il grafico bi-plot consente di visualizzare congiuntamente i coefficienti delle prime componenti principali ortonormali per ciascuna delle 44 variabili e i coefficienti delle prime componenti principali ortonormali (scores) per ciascuna delle 821 osservazioni (campioni).

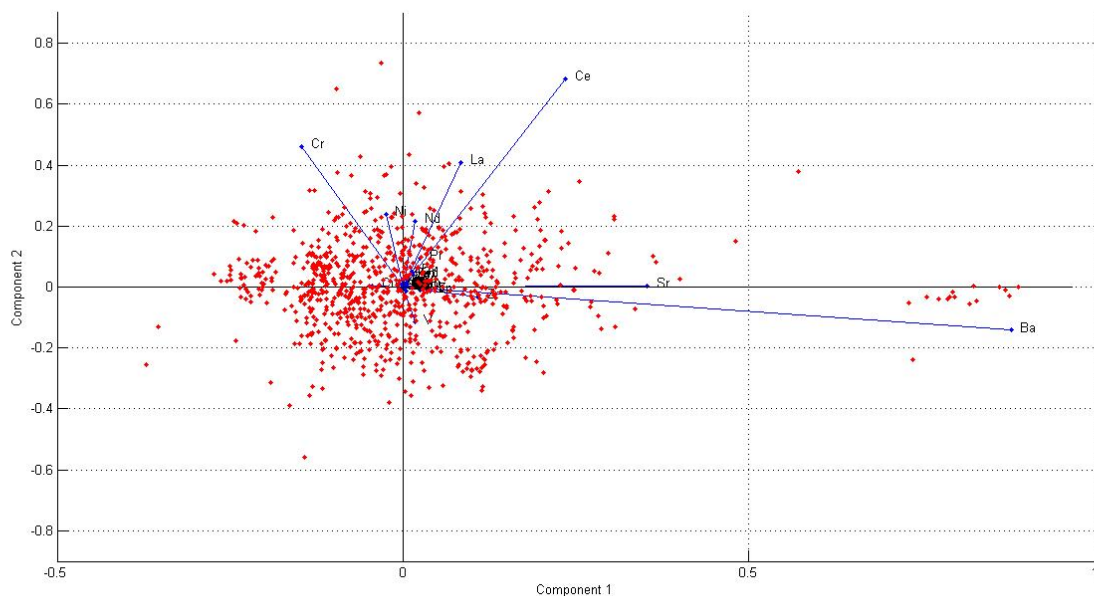


Figura 3: grafico bi-plot rispetto alle prime due componenti principali.

Ciascuna delle 44 variabili è rappresentata da un vettore la cui direzione e lunghezza quantificano il contributo della variabile alle componenti principali che costituiscono il nuovo sistema di riferimento [Figura 3].

3.2 Fase 1 - Risultati

La prima componente principale (asse orizzontale) mostra, per quanto visibili con il fattore di ingrandimento impostato, coefficienti positivi per gran parte delle variabili e cioè per Bario, Stronzio, Cerio, Lantanio, Neodimio (e altri) e coefficienti negativi per Cromo e Nichel (e altri). La seconda componente principale (asse verticale) mostra, per quanto visibili con il fattore di ingrandimento impostato, coefficienti positivi per Cerio, Stronzio, Lantanio, Cromo, Nichel, Neodimio (e altri) e coefficienti negativi per il Bario (e altri).

E pertanto, è stato possibile affermare che:

- I maggiori coefficienti per la prima componente principale corrispondono agli elementi Ba, Sr, Ce, La.
- La seconda componente principale consente di distinguere tra campioni caratterizzati da elevati valori di concentrazione di Ce, Sr, La, Cr, Ni, Nd (e altri) e campioni caratterizzati da piccoli valori di concentrazione di Ba (e altri).

3.3 Fase 2 - Scenario: intercettazione di campioni radioattivi, verifica della consistenza reciproca e verifica della consistenza con la NNFL

Il Gruppo di Lavoro ENEA ha acquisito le informazioni, diffuse dagli autori dell'esercizio internazionale attraverso il portale dedicato "Slack", relative ai tre campioni intercettati.

SAMPLE ID	53023	8615-OSF-1	1005
U wt%	79,52	79,3	75,90
Ti (ppm)	0,30	0,30	0,20366
Al (ppm)	17,69	20,68	13,57
Fe (ppm)	7,28	6,53	0,85
Ca (ppm)	12,97	12,64	0,76
Mg (ppm)	6,43	4,64	0,10
Mn (ppm)	0,19	0,16	0,02
K (ppm)	0,18	0,23	6,26
Na (ppm)	1,31	1,74	2,17
P (ppm)	0,06	0,07	0,03
La (ppm)	204,10	408,35	16,03
Ce (ppm)	299,91	600,07	33,43
Pr (ppm)	30,71	61,76	4,27
Nd (ppm)	86,83	173,54	17,58
Sm (ppm)	14,87	29,28	3,98
Eu (ppm)	0,37	0,85	1,30
Gd (ppm)	12,68	-	4,18
Tb (ppm)	2,46	5,21	0,72
Dy (ppm)	15,42	31,14	4,56
Ho (ppm)	3,48	6,54	0,97
Er (ppm)	11,71	24,05	2,80
Tm (ppm)	1,65	3,30	0,41
Yb (ppm)	10,35	20,82	2,53
Lu (ppm)	1,63	2,73	0,40
Sc (ppm)	45	-	-
V (ppm)	309	-	12
Cr (ppm)	143	-	246
Co (ppm)	35	-	0,8
Ni (ppm)	44	13	4,9
Cu (ppm)	92	-	325
Zn (ppm)	81	-	12
Ga (ppm)	13	-	18,35
Rb (ppm)	2,6	3,4	166
Sr (ppm)	358	692	166,5
Y (ppm)	14	13,9	13
Zr (ppm)	33	27	24,6
Nb (ppm)	0,25	0,2	19
Cs (ppm)	0,26	-	-
Ba (ppm)	87	107	-
Hf (ppm)	0,94	-	-
Ta (ppm)	0,02	-	-
Pb (ppm)	2,00	1,16	16,5
Th (ppm)	0,197	-	25,3
Si (ppm)	0,075	-	-

Tabella 4: composizione elementare dei tre campioni intercettati.

La valutazione della consistenza fra i tre campioni ignoti e fra i campioni ignoti e la NNFL, è stata portata a termine utilizzando un metodo basato sugli schemi delle Terre Rare (REE) come principali indicatori per l'appartenenza alle delle classi di affinità.

L'utilizzo dei "trace elements", come riportati nel database originario, è stato giustificabile poichè la variazione della loro concentrazione è connessa con la geochimica delle rocce e con le formazioni geologiche e fenomeni di degradazione in modo che essi possano essere utilizzati per valutare l'origine delle rocce e dei processi (naturali o antropogenici) che hanno subito. Inoltre, una variazione significativa della loro concentrazione, confrontata con le concentrazioni attese (in matrici simili), potrebbe essere rappresentativa dei processi di produzione. L'utilizzo dei "trace elements" è dunque importante per affrontare il problema dell'origine di campioni sconosciuti. In particolare, anche il gruppo delle "Terre Rare" (REE) può essere utilizzato per verificare l'origine dei campioni e dei processi che hanno subito, con particolare riferimento ai processi tecnologici e alle applicazioni tecnologiche.

Gli schemi delle REE sono stati costruiti utilizzando i loro valori normalizzati rispetto alla condrite carbonacea (secondo il database "Taylor & McLennan, 1985), escludendo lo Scandio a causa della sua assenza in due dei tre campioni intercettati.

Nella [Figura 4], gli schemi delle Terre Rare dei tre campioni intercettati (53023, 8615-OSF-1, 1005) sono confrontati con i medesimi schemi della NNFL, per ogni tipo di basalto.

Le barre d'errore rappresentano l'intervallo min-max per ciascun valore delle REE.

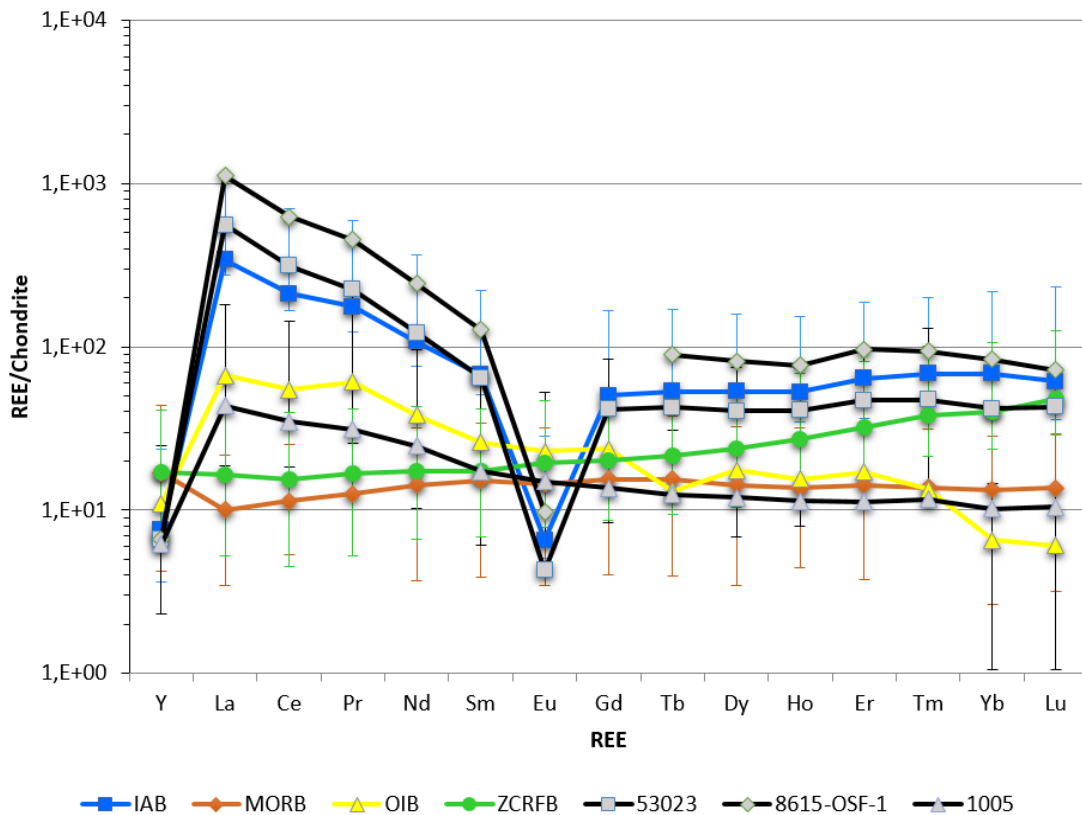



Figura 4: schemi delle REE dei tre campioni intercettati (punti in grigio), confrontati con gli schemi REE della NNFL (normalizzazione dal database "Taylor & McLennan", 1985).

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	12	49

3.4 Fase 2 - Risultati

Osservando la [Figura 4] è stato possibile rispondere ai primi due quesiti previsti per la Fase 2:

- Gli schemi REE dei primi due campioni intercettati (53023, 8615-OSF-1) si adattano bene agli schemi REE della classe di affinità del basalto IAB.
- Lo schema REE del terzo campione intercettato (1005) mostra un comportamento più ambiguo; esso, infatti, si adatta bene alla classe di affinità del basalto OIB fino all'Europio mentre dal Gadolinio al Lutezio sembra adattarsi meglio alla curva del basalto MORB.

L'inconsistenza reciproca fra i tre campioni intercettati rende più complesso lo scenario del traffico illecito suggerendo la possibilità di una infrazione multipla della catena di controllo dei materiali.

Riguardo al terzo quesito previsto per la Fase 2, sono state individuate, in un contesto prettamente scientifico/tecnico, le seguenti azioni principali:

- Condurre analisi radiologiche e chimiche sui pneumatici e sedili del veicolo intercettato e anche sugli indumenti degli individui a bordo.
- Selezionare un gruppo specializzato di esperti con abilità geologiche per investigare più approfonditamente le proprietà dei materiali intercettati.
- Condurre ulteriori analisi chimiche e test morfologici sui campioni intercettati per la deduzione delle origini geografiche e per rilevare qualsiasi indizio di processamento tecnologico implementato.
- Predisporre ispezioni nelle miniere di Uranio per controllare l'inventario e le licenze.


Si ritiene, inoltre, che, in un contesto di carattere investigativo, potrebbero risultare utili anche le azioni di seguito riportate:

- Allertare la IAEA e i paesi limitrofi.
- Ricostruire i movimenti del veicolo con il carico illecito.
- Fermare e controllare altri veicoli che appartengano alla stessa azienda di trasporti cui appartiene il veicolo intercettato.
- Incrementare i controlli ai confini attraversati dal veicolo con il carico illecito.

Riguardo al quarto quesito previsto per la Fase 2, al fine di condividere le informazioni tra le istituzioni governative, si potrebbe redigere un dettagliato rapporto tecnico confidenziale che sia distribuito alle competenti autorità pianificando riunioni di coordinamento dedicate a gruppi specializzati.

4 SOLUZIONE DELL'ESERCIZIO

Gli autori dell'esercizio hanno prodotto un rapporto per la Fase 1 e un rapporto per la Fase2, proponendo la propria soluzione per i primi due quesiti della Fase 2.

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	13	49

4.1 Fase1

Il rapporto prodotto dagli autori per la Fase 1 è un sommario delle risposte di tutti i gruppi che hanno partecipato all'esercizio.

Il rapporto è riportato in Appendice B.

4.2 Fase 2

Il rapporto prodotto dagli autori per la Fase 2 è un sommario delle risposte di tutti i gruppi che hanno partecipato all'esercizio e comprende la soluzione proposta dagli autori stessi riguardo ai quesiti 1 e 2.

Il rapporto è riportato in Appendice C.

5 Confronto tra le soluzioni

Si riporta di seguito il confronto tra le soluzioni.

5.1 Fase 1.

Sia gli autori dell'esercizio, sia il Gruppo di Lavoro ENEA hanno scelto di utilizzare un comune foglio di calcolo Microsoft Excel per realizzare la libreria. Tale scelta è in sintonia con la necessità di una veloce consultazione da parte di esperti con diverse competenze informatiche e di uno scambio di informazioni efficace, per cui l'utilizzo di un programma ben noto alla maggior parte degli esperti è condiviso da entrambi gli approcci. Tale scelta, d'altra parte, presenta alcune limitazioni in termini di funzionalità aggiuntive, rispetto ad altri programmi di archiviazioni più specifici, o laddove ci sia una numerosità e complessità superiore di informazioni riguardanti le sorgenti. Rimane in ogni caso fondamentale l'approfondimento della necessità di interscambio delle informazioni e di protezione dei dati sensibili.


5.2 Fase 2.

Il Gruppo di Lavoro ENEA, confrontando gli schemi REE del database e dei campioni intercettati, è giunto ai seguenti risultati:

1. I primi due campioni intercettati "53023" e "8615-OSF-1" si adattano bene agli schemi REE della classe di affinità del basalto IAB e sono pertanto reciprocamente consistenti.
2. Lo schema REE del terzo campione intercettato "1005" mostra un comportamento più ambiguo; esso, infatti, oltre a non essere consistente con gli altri due campioni intercettati, si adatta bene alla classe di affinità del basalto OIB fino all'Europio mentre dal Gadolinio al Lutezio sembra adattarsi meglio alla curva del basalto MORB [Figura 4].

Gli autori dell'esercizio hanno riportato i seguenti risultati:

1. I campioni "53023" e "8615-OSF-1" sono reciprocamente consistenti ed è identificata l'affinità IAB nel modello della NNFL.
2. Il campione "1005" è risultato non consistente con gli altri due campioni o con nessuna delle quattro classi di affinità rappresentate nel modello della NNFL.


	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	14	49

6 CONCLUSIONI

Dal confronto tra le soluzioni per la Fase 2 si evince quanto segue:

1. La soluzione degli autori è stata coerente con la soluzione del Gruppo di Lavoro ENEA (GLE): i campioni intercettati “53023” e “8615-OSF-1” risultano reciprocamente consistenti ed entrambi consistenti con la classe di affinità IAB del modello della NNFL.
2. La soluzione degli autori è stata parzialmente coerente con la soluzione trovata dal Gruppo di Lavoro ENEA. Entrambi i gruppi hanno concordato sulla non-consistenza del campione “1005” con gli altri due. Riguardo alla consistenza del campione “1005” con le classi di affinità del modello della NNFL, il GLE ha letto lo schema REE attribuendo al campione un adattamento parziale e simultaneo alle REE delle classi di affinità IOB e MORB. Gli autori hanno invece affermato la non-consistenza del campione in oggetto con alcuna delle classi di affinità.

Gli autori non hanno fornito indicazioni sul metodo utilizzato per ottenere i propri risultati e pertanto la divergenza di cui al punto 2 potrebbe essere considerata meramente interpretativa anziché sostanziale.

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	15	49

7 ALLEGATO A: dettagli statistici Fase 1.

Descrizione sintetica della tecnica statistica “Analisi delle Componenti Principali” (ACP). [7] [8]

La formulazione più nota della ACP è attualmente quella proposta da Hotelling nel 1933 e si basa sull'ipotesi che i valori di un insieme di p variabili originarie correlate siano determinati da un più ristretto insieme di variabili mutuamente indipendenti. Tali variabili vengono determinate come combinazione lineare delle variabili originarie in modo tale da massimizzare il loro successivo contributo alla varianza totale dell'insieme. Il sistema delle osservazioni è organizzato in forma matriciale, costruendo una matrice pxq (righe x colonne), con q osservazioni e p variabili (mutuamente dipendenti ossia correlate). Le q osservazioni sono rappresentabili da punti in uno spazio p -dimensionale (spazio delle righe).

Obiettivo della ACP è ridurre la dimensionalità dello spazio delle righe, in modo da rappresentare la variabilità del problema utilizzando un minor numero di fattori, ossia utilizzando il minor numero di dimensioni possibile (dimensionalità intrinseca del problema). Si tratta, dunque, di costruire un sottospazio m -dimensionale ($m < p$) i cui assi coordinati (le componenti principali) si allineino con l'insieme dei dati originari.

Le p variabili correlate originarie, costituiscono una variabile multivariata rappresentata da un vettore in uno spazio p -dimensionale. Per tale variabile multivariata è possibile definire una funzione densità di probabilità (FDP) multivariata, con i propri parametri di localizzazione e dispersione e cioè un vettore valor medio e una matrice di covarianza. Quest'ultima quantifica l'ampiezza della FDP definendo il grado di correlazione tra le p variabili originarie ed è una matrice non-diagonale.

La ACP consente di diagonalizzare la matrice di covarianza, per mezzo di una opportuna variazione del sistema di riferimento (rotazione), ottenendo così un nuovo insieme di variabili ortogonali mutuamente indipendenti e di conseguenza la FDP multivariata è esprimibile come una semplice produttoria di FDP univariate. In tal modo le nuove variabili (Componenti Principali) sono combinazioni lineari delle variabili originarie e poiché la matrice di covarianza nella base delle Componenti Principali è diagonale, la varianza totale è la somma delle singole varianze delle Componenti Principali.

In sostanza, lo scopo della ACP è la rappresentazione di un insieme di dati in uno spazio p -dimensionale, con matrice di covarianza non-diagonale, in un sottospazio m -dimensionale ($m < p$) con matrice di covarianza diagonale. La ACP è classificata come metodo del secondo ordine poiché la riduzione della dimensionalità dello spazio è basata unicamente sulle proprietà della matrice di covarianza (momento del secondo ordine).

La prima Componente Principale, ossia il primo asse del nuovo sistema di riferimento nel sottospazio m -dimensionale, è calcolata in modo tale da descrivere una quantità di varianza originaria maggiore della varianza associata a ciascuna singola variabile. In altri termini la prima Componente Principale è la direzione in cui si realizza la massima dispersione dei dati (criterio su cui si basa la formulazione di Hotelling) e quindi essa spiega la massima percentuale della varianza dei dati, in una dimensione.

Calcolando la seconda Componente Principale si otterrà la direzione ortogonale in cui si realizza la massima dispersione dei dati, ma inferiore alla precedente, e pertanto le due componenti principali spiegheranno la massima percentuale della varianza originaria in due dimensioni. E così via, le componenti principali successive spiegheranno una sempre minor percentuale della varianza originaria, nel sottospazio m -dimensionale. La conoscenza della

percentuale di varianza spiegata è essenziale per l'interpretazione dei grafici delle componenti principali.

In estrema sintesi formale si ha che, date p variabili osservate X , si costruiscono p combinazioni lineari Y del tipo:

$$\begin{aligned}
 Y_1 &= a_{11}X_1 + a_{12}X_2 + \dots + a_{1i}X_i + \dots + a_{1p}X_p \\
 &\quad \dots \\
 Y_k &= a_{k1}X_1 + a_{k2}X_2 + \dots + a_{ki}X_i + \dots + a_{kp}X_p \\
 &\quad \dots \\
 Y_p &= a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pi}X_i + \dots + a_{pp}X_p
 \end{aligned}$$

Le combinazioni lineari Y , nominate "Componenti Principali", devono soddisfare le seguenti condizioni:

1. Assenza di correlazione:

$$\text{Cov}(Y_{k'}, Y_k) = 0, \forall k' \neq k$$


2. Ordine in relazione alla quantità di varianza complessiva che ciascuna variabile può sintetizzare:

$$V(Y_1) \geq V(Y_2) \geq \dots \geq V(Y_i) \geq \dots \geq V(Y_p)$$

3. La varianza complessiva dei due sistemi di riferimento deve coincidere:

$$\sum_{j=1}^p V(Y_j) = \sum_{i=1}^p V(X_i)$$

Il sistema delle X_i variabili originarie è ridondante essendo esse correlate. Di conseguenza, per riassumere adeguatamente i dati, sarà sufficiente considerare solo le prime m ($m < p$) componenti principali, ammesso che queste forniscano un adeguato contenuto informativo. Il criterio più immediato per la scelta del numero m di componenti rispetto a cui rappresentare il fenomeno presuppone che venga prefissata la proporzione della variabilità complessiva che si desidera tali componenti spieghino, ad esempio i due terzi oppure il 70/80%; m sarà perciò il minor numero di componenti per le quali tale percentuale è superata.

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. di 17 49
--	--	------------------	----------------------	-------------------------

8 ALLEGATO B: Summary Report on GSv3 – Phase 1




Galaxy Serpent Version 3 (GSv3)



Summary report on Galaxy Serpent version 3 – Phase 1

Contents

- Explanatory Notes.....Page 18
- Summary of Responses to Question 1.....Page 19
- Summary of Responses to Question 2.....Page 20
- Summary of Responses to Question 3.....Page 21
- Summary of Responses to Question 4.....Page 22
- Summary of Responses to Question 5.....Page 23
- Sampling of Submitted Figures.....Page 26
- Acronyms Used.....Page 35

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	18	49


Explanatory Notes

The following report is a summary of responses from teams submitting Phase 1 reports for the Galaxy Serpent – version 3 exercise. 29 teams comprising 132 individuals participated in the GSv3 exercise. 22 of these 29 teams (76%) submitted Phase 1 reports as of December 15, 2017. Responses provided by teams are summarized by question prompt below. Results cited reflect the work of teams answering a given question. Not all reports addressed each question, and multiple tools (Excel, R, etc.) or statistical methodologies were utilized by teams, thus the numbers cited do not necessarily sum to the number of reports submitted. Some teams also submitted free-form reports rather than responding question-by-question, and on occasion reports did not address questions directly – in such cases best inferences or judgements were made. Individual quotes, graphs, or results are given without attribution to maintain anonymity, and serve to illustrate some of the various approaches and methodologies teams employed.

Acronyms used are identified on the final page of this document. About 8 of the 22 respondents included plots or graphs in their reports. Some of these are included near the end of this report. They were not selected on the basis of efficacy or to suggest more frequent use by teams, but rather to present a representative sample of different techniques that teams reported utilizing in organizing the data and looking for discriminating signatures.

This document is **not** intended as an “answer key” or set of “how-to” instructions, nor to convey how the exercise organizers “expected” teams to proceed through Phase 1. Rather, its purpose is simply to report the techniques various teams applied in terms of approaches and methodologies in organizing the provided data, which sometimes led to different preferred analytical paths. For example, some teams viewed trace elements as the main discriminator (largely due to gaps in data provided for REEs), while others addressed the missing REE data and viewed them as the main determinant.

The Galaxy Serpent exercise organizers would again like to thank all teams and individual participants for their dedication and involvement, and to acknowledge the devoted effort by all. We recognize that this work is conducted in addition to your normal duties, and deeply appreciate the commitment and effort both in this exercise, and in maturing the discipline of nuclear forensics, of all those involved.

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 19	di 49
--	--	------------------	----------------------	-------------------	-----------------

Summary of Responses to Question 1

- 1) Prompt: *How did you organize the data? Did you use all of the signatures provided? If not, which ones did you use or not use, and why?*


There was no universal approach employed by teams when organizing the data but many common themes emerged. A variety of valid methods were used, with many teams combining multiple approaches. Teams generally ignored the latitude/longitude fields as they were unrelated for material discrimination, and some retained the sample ID while others did not.

All teams kept the “affinity” or “class” descriptor (OIB, MORB, IAB, or ZCRFB), with many making separate worksheets for each. Teams employed a variety of techniques to categorize the data, ranging from elementary to more complex in nature, and looked for elemental groupings providing distinctive signatures. Teams generally combined machine learning approaches, using algorithms that can learn from data without relying on rules-based programming, with statistical or latent variables approaches, looking for discerning relationships among given or processed data. Reports varied in the level of specificity regarding the methods employed, but a couple teams stated that they used box plots and/or histograms to investigate which elements and/or groupings of elements might prove useful for distinguishing affinities. A couple teams reported using histograms to combine REE data (especially Eu, Tb, Yb, Lu) with Ti, Zr, Hf, Nb, Ta (group IV and V) data; and combining group IV transition element Zr, Hf, Ti data with alkaline-earth/group IIA elements (Ca, Sr, Ba) and Pb data.

All teams seem to have separated data by affinity, and most additionally divided data along the lines of trace elements and REEs. For REEs, some teams used both provided data and “normalized to chondrite” values, while the majority discussed using only the normalized values. One team reported challenges in finding all normalization factors. Another found the “analysis of REE normalized pattern is a very powerful technique for the identification of UOC origin,” and another also decided to “focus on lanthanides in the dataset, as they are a good indicator of UOC-origin.” One team noted their decision to use REEs (after chondrite normalization) and additional key elements like Ti, Mn, K, Th, Zr, or elemental correlations such as Ca/Sr, Na/K. Most teams did not go into this level of specificity in their reports, but many seemed to have come to thematically similar conclusions.

At least one team assessed data based on discerning geologic ratios (eg. La/Sm vs Zr; Nb/Y vs Zr). One team described separating data using classifications of small vs large ionic radii elements and REEs (received, and chondrite normalized), which were analyzed to evaluate diagnostic properties, and key correlations and covariances.

The primary difference in how teams organized data centered on how they handled the missing data. About 5 teams explicitly discussed the possible approaches they saw in this regard, noting that one could i) replace missing data by mean or median values, which could affect reliability; ii) use dummy variables to fill gaps, which some noted may not be the best option for a large numerical dataset; or iii) delete the deficient data, which may

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 20	di 49
--	--	------------------	----------------------	-------------------	-----------------

be adverse for multivariate statistics as it can artificially reduce data. Among teams who specifically noted their approach in this regard:

- Three teams chose to ignore elemental data that was over 30% deficient.
 - One of these teams noted that for this reason they did not use Fe, Sc, Co, Cu, Zn, Ga, Nb, Sn, Cs, Hf, Ta, Pb, Th, Si, I, Sn, or the lanthanides due to effect missing data would have on reliability
- One team noted excluding Fe, Sn, Ga due to the sparse nature of data
- One team reported they considered all elemental data, but excluded those with missing values on a case-by-case basis by considering the effect of these missing values
- Two teams opted to exclude U as, one team stated, “it explains 95% of the variance (a very biased result)” and because “it is a main component (not a trace element), and thus not useful in a traceability investigation”
- Six teams opted to use “all signatures”


Summary of Responses to Question 2

- 2) Prompt: *Which technical approaches did you use to evaluate and interrogate the provided UOC data? (For example, statistical regression, principal component analysis, comparative/correlative assessments, etc.)*

Teams varied in the degree of detail given in outlining the technical approaches they employed. A few teams used fundamental Excel analysis as their primary tool, while the majority of teams used Excel as a precursor to discern discriminatory patterns and/or prepare given or processed datasets for more complex analysis. Teams using Excel noted looking at ranges, histograms, box plots and the like. Two teams noted using ANOVA and descriptive statistics prior to additional analysis.

Techniques teams reported using for further analysis included both latent variable and machine learning approaches, with most teams using multiple methods in combination. Due to the varying degrees of specificity in discussing the specific methods used by teams, it was sometimes challenging to clearly characterize the methodologies used. With this stipulation, it appears that, of the 22 teams submitting reports:

- Fifteen teams used PCA
- Four teams using PCA did so in conjunction with PLS-DA, “a machine learning algorithm that attempts to explain variance through latent variables while simultaneously dividing data into classes, with imputation being used to fill in the dataset due to the large amount of missing data”
- Four team reported using PCA in conjunction with OPLS-DA
- Four teams reported using “statistical/multivariate analysis methodologies” including cluster analysis with PCA; LDA for outlier testing and comparative assessment; and MVA based on a self-developed algorithm
- Two teams used “discriminant analysis”
- Three teams employed radar or spider plots to visualize data in their analyses

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 21	di 49
--	--	------------------	----------------------	-------------------	-----------------

- Four teams reported using classification/decision trees, random forest ensembles, and neural networks for comparative assessments of impurity distributions, ratios and chondrite-normalized REE patterns
- Three teams reported using “statistical regression”
- One team noted using “hierarchical cluster analysis”
- One team noted using “pattern analysis”
- One team reported using REE data for a “consistency check” for each affinity, looking for outliers, calculating relative differences, and plotting selected data
- One team reported using “comparative and correlative methods”
- Two teams noted using “comparative assessments using elemental ratios” by normalizing all data to standard chondrite values, and then plotting the normalized concentrations by increasing atomic number
- One team reported using “subjective and variational-statistical estimation of data, as well as graphical visualization of data sets”

PCA was by far the most commonly named technical approach, being cited explicitly by at least 15 of the 22 teams submitting Phase 1 reports. Teams generally applied PCA using trace element concentrations to observe data clustering. Some teams reported poor discrimination for PCA applied to REEs due to missing data, while other teams who replaced missing data (see page 3) reported being able to reduce dimensionality. One such team noted they found the first three principal components explained 85% of variance, allowing them to reduce dimensionality to 3. Another team reported using PCA as “unguided learning technique to look for groups/patterns in data” coupled with discriminant analysis to improve the discrimination ability between perceived groups.

Teams analyzing REE data reported a series of possible paths, sometimes used in conjunction, including i) before normalization without replacing missing data; ii) after chondrite normalization without replacing missing data; iii) before normalization replacing missing data using geometric means/medians; and iv) after chondrite normalization replacing missing data using geometric means/medians. One team reported determining geometric means using an ALS algorithm.


Summary of Responses to Question 3

- 3) Prompt: *Did you use a commercial software package to evaluate and interrogate the data? If so, which one? (For example, Excel, R, Sigma, etc.)*

Most, but not all, teams utilized Microsoft Excel, many citing its availability and convenience since the data was provided as .xls files. While not always explicitly state it appears that, of the 22 respondents, roughly:

- 4 (~18%) used only Excel
- 15 (~68%) used Excel in concert with other software packages
- 3 (~14%) did not utilize Excel in any significant way

The majority of teams seem to have used some combination of free (R, Simca-P, etc.) and commercial (Matlab, Origin, etc.) software packages. Of the teams using Excel in combination with other software packages, Excel was often utilized for some lower-level tasks (data preparation, initial data evaluation, preliminary analysis and comparison, REE

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	22	49

normalization, etc.) while additional software packages were used for more complex analyses. Teams reported using the following non-Excel software packages:


- R; a free package (6 teams)
 - One team additionally used Shiny (R package for interactive web application)
 - Teams used R in statistical analysis environment for PCA, LDA decision trees and/or randomForest analyses
- Matlab; a commercial package (4 teams)
 - For MVA, PCA, and/or developing homemade algorithms)
- Origin/OriginPro; a commercial package (3 teams)
 - For histograms, boxplots, ANOVA, PCA
- Simca-P; a free package (2 teams)
- SPSS Statistics; a commercial package (1 team)
- Anaconda; a free package (1 team)
 - Used with Orange, a data mining add-on
- PAST; a free package (2 teams)
- JMP; a commercial package (1 team)
- Solo from Eigenvector Research; a commercial package (1 team)
- Python; a free package (1 team)
 - Planned use for future (post-exercise) analysis of provided data

Summary of Responses to Question 4

4) Prompt: *What was your approach if you did not use commercial software? Did you develop your own tool, or did you take a different approach? (For example, subject matter expert-driven)*

The 22 responding teams interpreted this question in a variety of ways. 12 explicitly noted the essential role that SMEs played in driving the analysis, while 10 cited no other “tools” being used. However, in the latter group SME-driven analysis clearly played a vital and necessary role in working through Phase 1 of the exercise. Those explicitly citing the role of SMEs used their proficiency in the following reported ways:

- SME-driven approach to signature selection, data analysis and interpretation (including REE interpretation and pattern analysis); drawing of confidence intervals (7 teams)
- SME developed indigenous code (5 teams)
 - For use in Phase 2 to help relate unknown material data to specific samples from the model NF database (1 team)
 - For developing an MVA original computation tool for material discrimination analysis use with NF in Phase 2 to help compare unknown material data to specific samples from Phase 1 database (1 team)
 - For developing a homemade specific statistical tool (3 teams)
- Several teams alluded to the vital role SMEs play, but noted that this would not be fully leveraged until the Phase 2 analysis

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 23	di 49
--	--	------------------	----------------------	-------------------	-----------------

Summary: While not explicitly mentioned in the responses, all teams clearly used a SME-driven approach (whether in statistics, geochemistry, nuclear forensics, etc.) to data analysis and interpretation, with several teams alluding to the critical role this will play in Phase 2. Teams also leveraged SMEs in developing indigenous code for handling and analyzing the data.

Summary of Responses to Question 5

5) Prompt: *What features of the provided data were the most useful?*

The vast majority of respondents did not address this lead-in question, instead placing their responses in the sub-questions below. For those that did respond:

- One team reported trace elements/impurities were the most useful data
- One team reported that no single variable was “most useful,” but the fact that all four classes contained enough information to give distinct patterns was of the greatest discriminatory benefit
- One team reported the process of interrogating data graphically with histograms and box plots comparing data for the four affinities led to their conclusion that REE and select trace element concentrations were the most useful.

a. Prompt: *Did you utilize isotopics? Why or why not?*


All teams noted that uranium isotopic concentrations were not provided, thus isotopics were not used. Several teams used ratios of various provided elemental concentrations as a discriminating signature. Some teams added that some useful isotopic ratios that could be key NF signatures, if isotopic concentrations had been provided, would be ($^{234}\text{U}/^{238}\text{U}$), ($^{235}\text{U}/^{238}\text{U}$), and Sr ($^{87}\text{Sr}/^{86}\text{Sr}$), and Pb ($^{208}\text{Pb}/^{204}\text{Pb}$) ratios. Some teams also noted that the $^{234}\text{U}/^{230}\text{Th}$ ratio can be used to calculate UOC age assuming all Th was removed from the sample during cleaning, and that any loss or addition of Th ($^{228}\text{Th}/^{232}\text{Th}$) could also be used as an isotope chronometer.

b. Prompt: *Did you utilize trace elements? Why or why not?*

All teams used trace element concentrations with the exception of one which reported that the missing data limited reliability as a classification/identification tool, but thought they might be useful when comparing to an unknown in Phase 2.

Several of the teams using trace elements reported that they “are one of the most significant signatures for the material discrimination analysis.” About 10 teams specifically noted their providing a “useful signature for clues about processing history and geolocation” as well as the potential manufacturer. In discussing the potential for origin assessment, several teams explicitly noted that trace elements can “indicate special technological processes (added chemicals or contaminants) which can identify/exclude producers.”

In terms of specific analytical methods, many reiterated that they used both trace element and REE signatures in combination. Additionally, teams noted the usefulness of the following techniques in considering trace elements:

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 24	di 49
--	--	------------------	----------------------	-------------------	-----------------

- Three teams noting organizing trace elements for each affinity into five levels based on concentration, and some teams stated they constituted the main useful data signatures in their analyses
- One team noted using box plots and histograms to assess all trace element data, and then focusing further analysis to a select few trace elements based on their variance
- One team cited observed correlations between trace elements for different affinities
- One team noted examining trace element data, but not finding it very useful at this point
- One team discussed limited success in using trace elements arising from challenges in finding normalization factors for some elements, and there being missing data for others, finding that elements (Ba, Ta, and Nb) and a plot of normalized concentration ratios of Ba/Nb vs Ta/Nb appeared distinctly unique across the four different affinities and could aid as a significant signature

Also, as noted earlier in the Question #2 discussion:

- Several teams again mentioned PCA as a valuable technique
- One team mentioned machine learning techniques using SME knowledge
- One team noted they split the data into three groups based on i) REEs, and ii) small and iii) large atomic radii to help identify the most diagnostically valuable elements and groupings

c. Prompt: *Did you utilize rare earth elements? Why or why not?*

All teams responding reported using REE data which many teams stated had strong discriminating power as seen by the differences in their mean concentrations across the four affinities. Many teams wrestled with the question of whether the high discriminating power of the REEs was overshadowed by the large amounts of missing data. For example, one team concluded that only Sc and Y had enough data to be included when considered along with trace elements but may have some value when comparing groups to “unknowns,” while another team decided only the lanthanides had enough data to be useful as an identifier regarding composition and potential origin.

Teams utilized REE data in multiple analyses. Several commented that there were fairly clear differences in REEs among the affinities, and several noted the particularly “marked Eu dip” in the IAB data which was not present for the other affinities. A couple teams stated that the lack of data for these elements mitigated their usefulness. One team succinctly summarized: “This raises the question: Is consistency within an affinity significant if there is only 6% of REE data present for any one affinity (OIB)?”

- Eight teams explicitly noted that REE signatures can also characterize local geochemical mineralogical conditions and features at the uranium mining site, and that samples of UOC from different places of origin will likely have individual

REE-signatures which can be linked to ore genesis and the geological origin and technological processing of the sample.

- In addition to noting that REEs constituted a very powerful technique for identifying UOC origin, one team noted that SEM images and XRD data would have been very useful attributive data for UOC characterization
- One team noted REE patterns are useful in determining geographical origin, but due to the sparsity of data used them as an auxiliary determinant
- Six teams discussed challenges in using REEs due to the sparsity of the data:
 - Due to the large amount of missing REE data, one team built a separate and smaller model which was “solely based on REEs with 112 samples (included any sample with at least one of 14 REEs included)”
 - One team normalized, and removed samples without REE data giving “useful dimensionality reduction”
 - One team went into some detail on the question of the impact of the significant amount of missing REE data, noting that “the influence of the REEs may be questionable.” This team made PCA scatter bi-plots (one of which is included for illustration in the figures) which conveyed the “discriminating power of the elements by their length and direction.” They noted that since “the majority of the REE point in a very similar direction..[t]his suggests that they all behave similarly when subjected to PCA analysis and therefore having data for all of the REEs may not be necessarily valuable for this method. Eu however points in the opposite direction to the other REE and may be more useful for PCA than the other REEs. The poor separation achieved...is likely attributed to the absence of the more discriminating REEs.”
- One team stated they used the REEs, but with limited success given their challenges with the randomly missing data, and in finding normalization factors for some elements. As referenced earlier, this same team found certain elements (Ba, Ta, and Nb) and derived plots of normalized concentration ratios of Ba/Nb vs Ta/Nb, appear “distinctly unique against the four different affinities and can aid as a significant signature”
- About 8 teams commented on aspects of the missing REE data as it related to effectively applying PCA.
 - One team used box plots and histograms of all REEs, a comparison of normalized REE data, and PCA to show that REEs were useful in distinguishing between affinities
 - One team reported using REEs by plotting chondrite normalized concentrations, and applying PCA. Based on population differences between classes, they developed a short list of trace elements for further PCA investigation, excluding others with data missing from further consideration in their analysis
 - One team noted that, “despite being sparsest dataset, REEs remain diagnostically valuable in the classification of UOCs” as PCA of REEs “showed

clearly defined clusters, and use of the chondrite-normalized REE patterns provides comparative distinction between the four affinities”

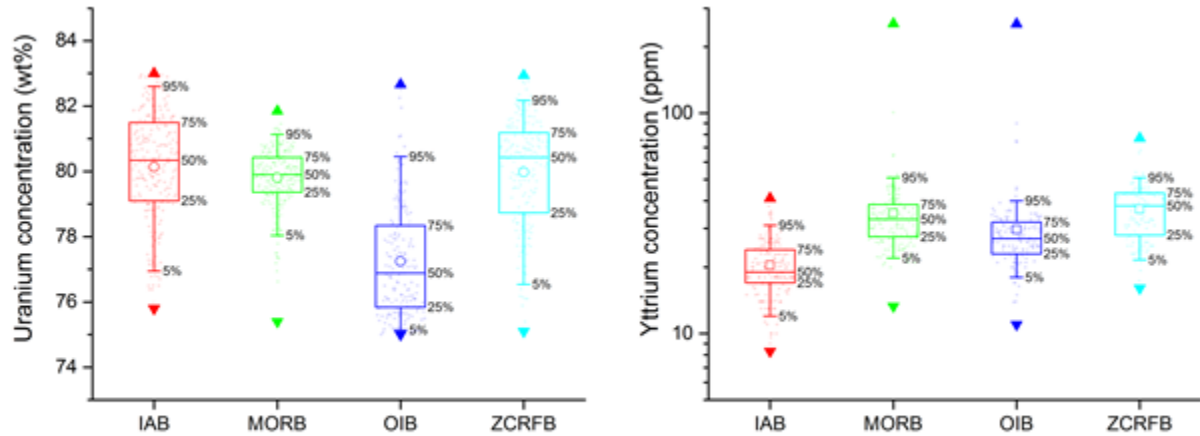
- One team noted that the analysis of chondrite-normalized REE content made it possible to separate the samples from the deposits. Also, analysis of graphs of the ratio of elements, such as Yb/Eu or Lu/Zr, made it possible to “practically guarantee that the sample belongs to one of four deposits,” and thus serve as a valuable discriminator
- One team used REEs in latent variables approaches utilizing SME knowledge

Sampling of Submitted Figures ¹

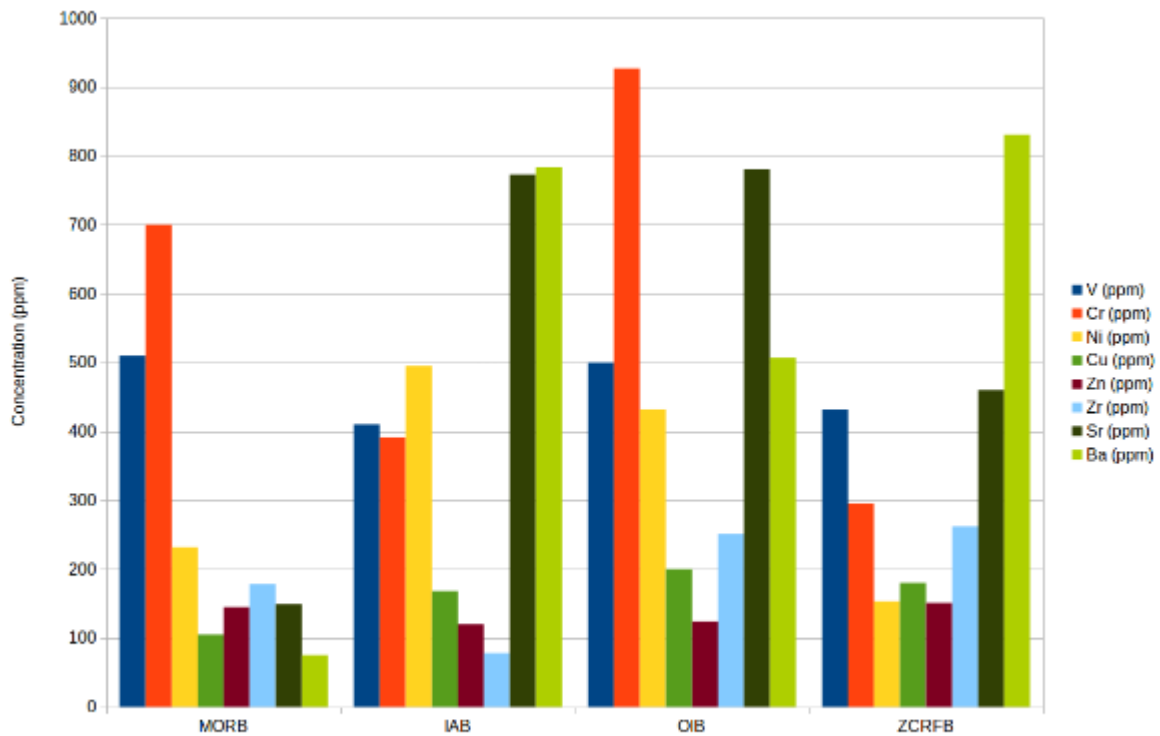
Table 1.2.1: Percentage of data present for each element in each Known Location

AFFINITY	U	Ti	Al	Fe	Ca	Mg	Mn	K	Na	P	La	Ce	Pr	Nd	Sm
IAB	100%	95%	94%	50%	94%	96%	94%	96%	94%	95%	9%	9%	9%	9%	9%
MORB	100%	100%	100%	63%	100%	100%	97%	100%	100%	95%	7%	6%	5%	15%	15%
OIB	100%	99%	99%	44%	99%	99%	98%	99%	99%	99%	9%	9%	8%	10%	10%
ZCRFB	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	21%	21%	21%	21%	21%
AFFINITY	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu	Sc	V	Cr	Co	Ni	Cu
IAB	9%	9%	9%	9%	9%	9%	9%	9%	9%	50%	66%	82%	37%	88%	52%
MORB	14%	6%	6%	6%	5%	5%	5%	7%	6%	71%	85%	87%	74%	90%	47%
OIB	9%	8%	8%	8%	8%	8%	6%	10%	9%	59%	70%	91%	49%	89%	62%
ZCRFB	21%	21%	21%	21%	21%	21%	21%	21%	21%	94%	63%	91%	56%	82%	74%
AFFINITY	Zn	Ga	Rb	Sr	Y	Zr	Nb	Sn	Cs	Ba	Hf	Ta	Pb	Th	Si
IAB	64%	37%	92%	96%	83%	88%	72%	5%	31%	92%	44%	38%	56%	63%	47%
MORB	50%	18%	82%	93%	72%	78%	60%	0%	28%	63%	33%	26%	27%	28%	13%
OIB	63%	32%	94%	95%	85%	86%	67%	2%	20%	94%	50%	33%	42%	69%	36%
ZCRFB	75%	47%	85%	88%	69%	84%	69%	0%	34%	100%	59%	50%	49%	91%	59%

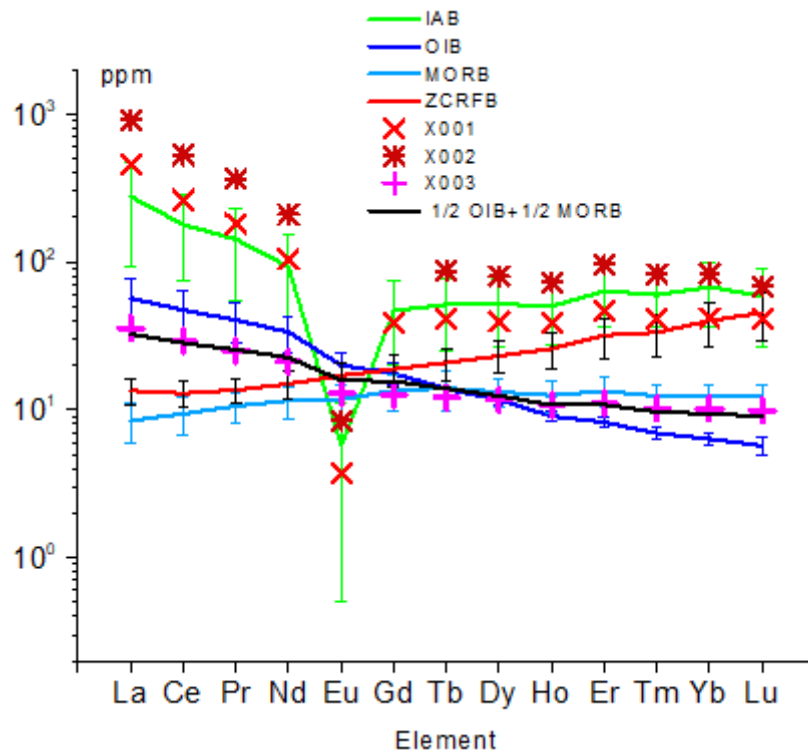
¹ *As noted in the introductory overview, about 8 of the 22 respondents included plots or graphs in their reports. Some of these are included near the end of this report. They were not selected on the basis of efficacy or to suggest more frequent use by teams, but rather to present a representative sample of different techniques that teams reported utilizing in organizing the data and looking for discriminating signatures.



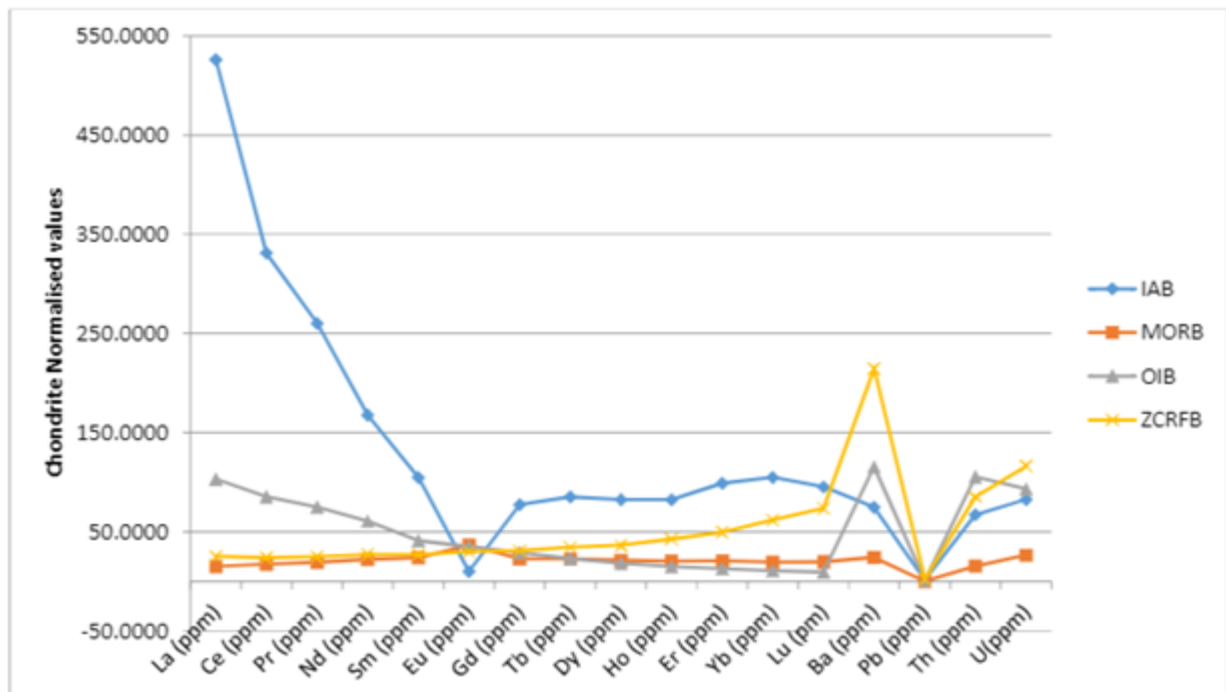
Box plots of elements showing a low level of discrimination among different affinity types (e.g., uranium and yttrium).

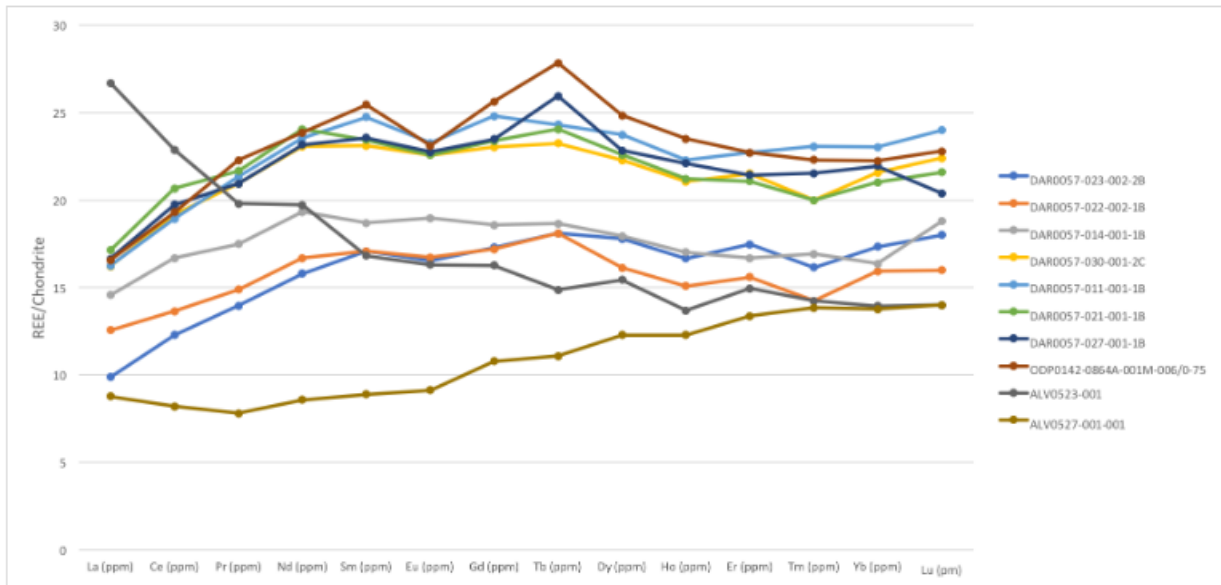


Concentrations of major trace elements in samples containing REE

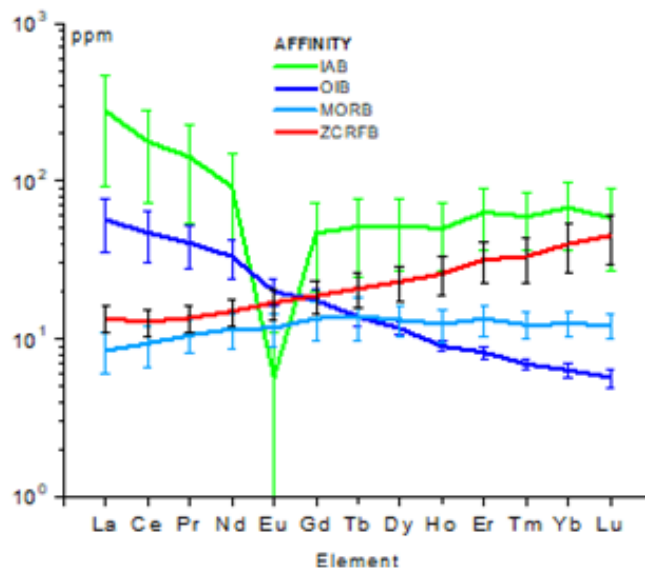


IAB, MORB, OIB and ZCRFB Chondrite Normalized averaged concentration values patterns

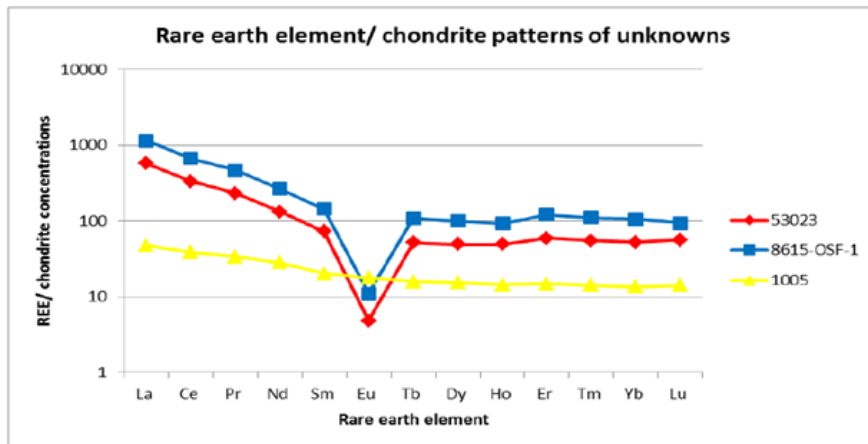




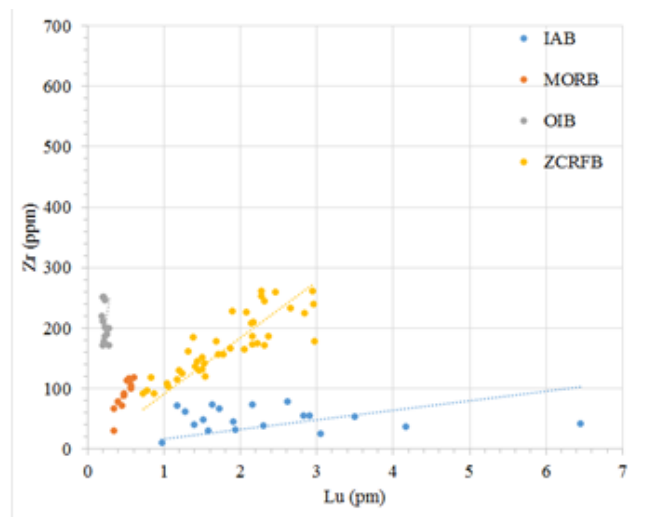
REE patterns for MORB category

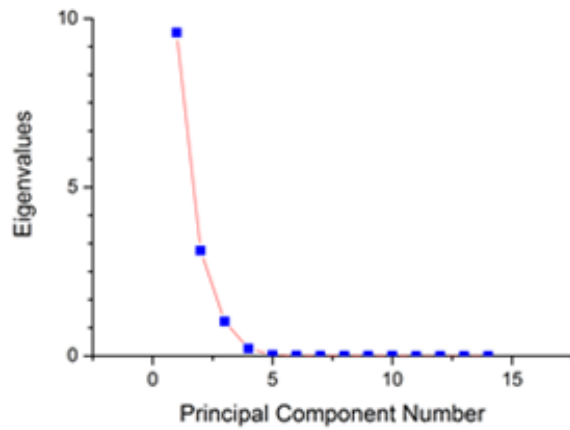


REE chondrite-normalized pattern for different AFFINITY of the samples in our database (software package ORIGIN-2017)

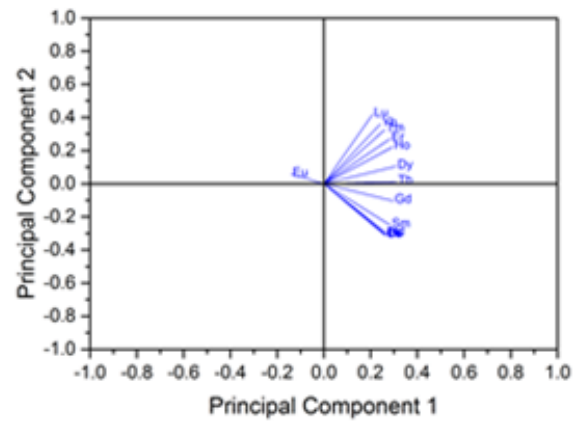


Chondrite normalised REE patterns for the unknowns

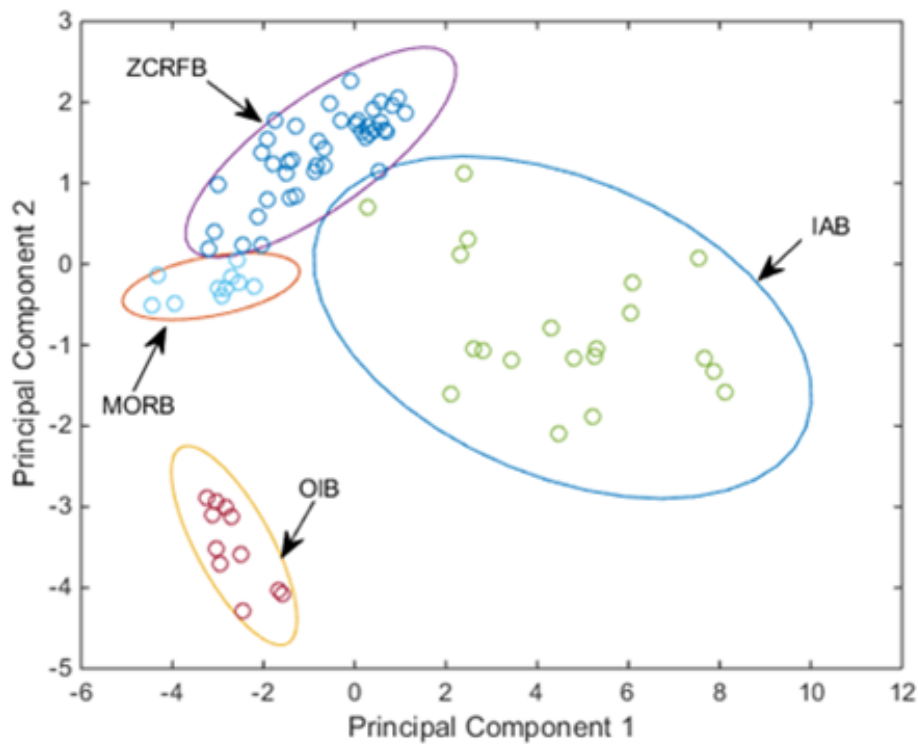




(a)

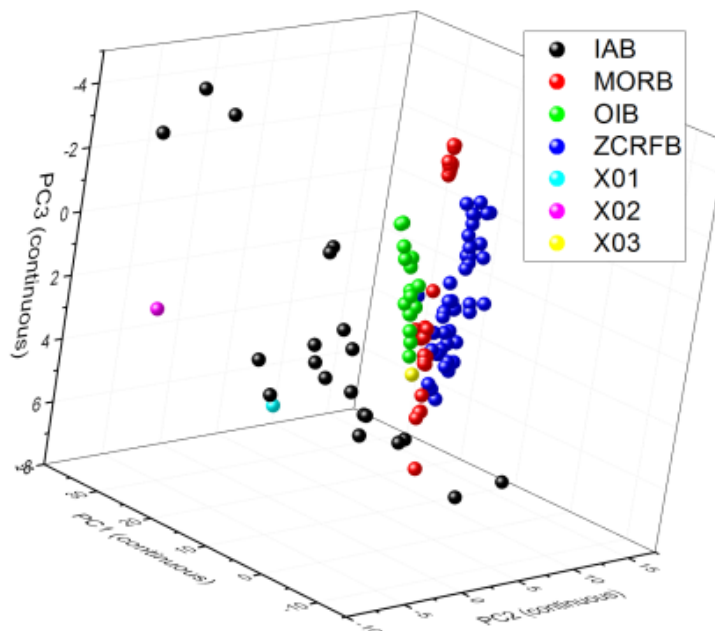
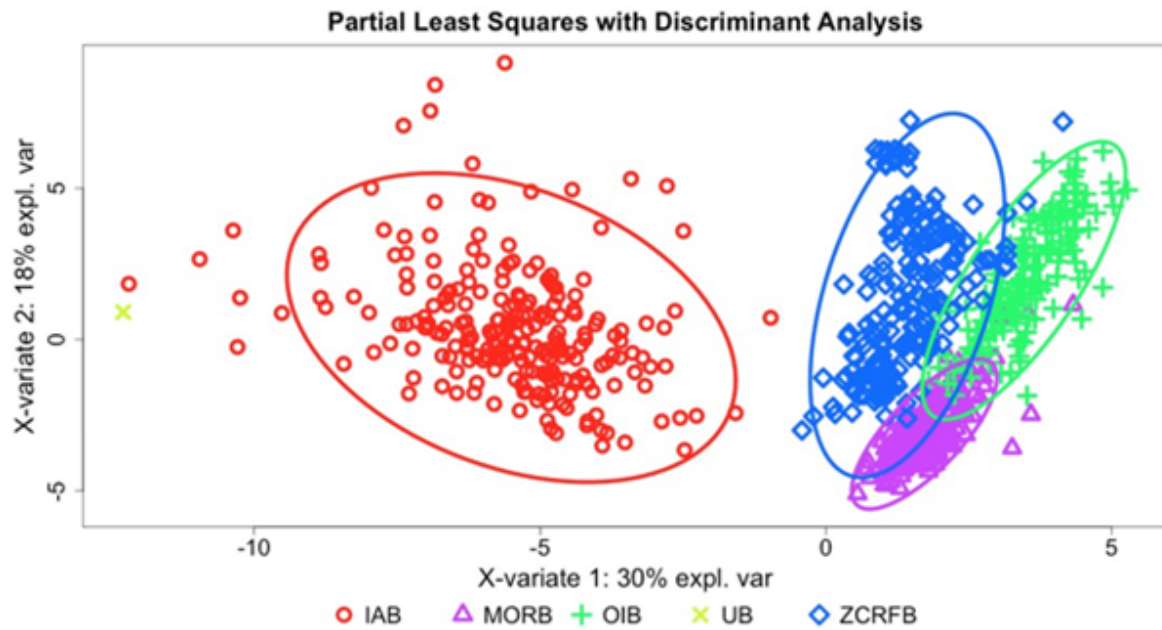


(b)

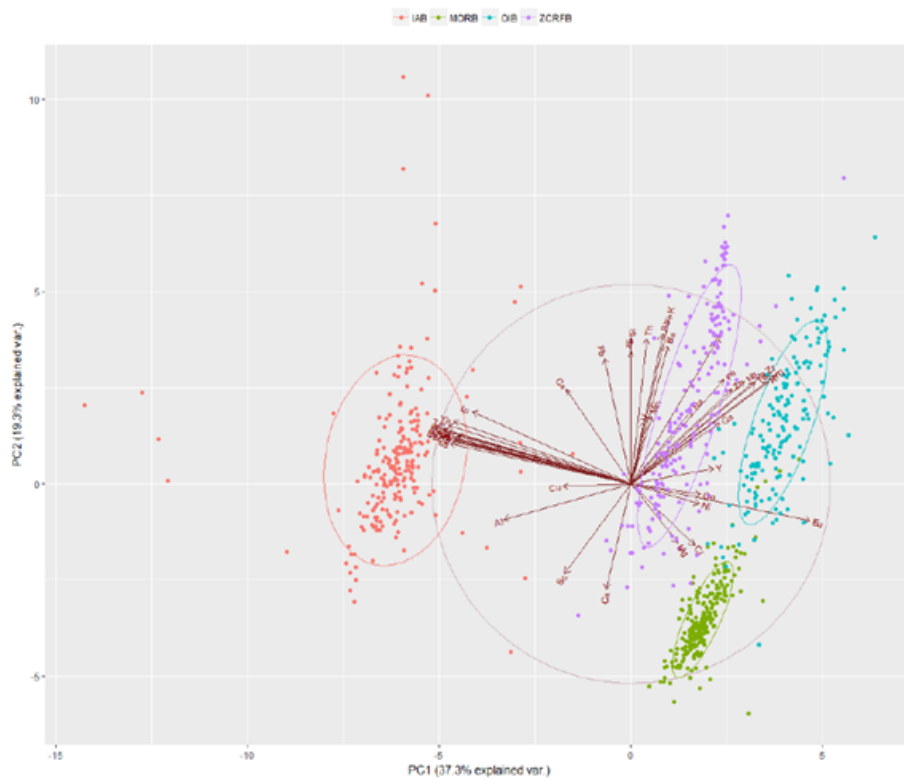


(c)

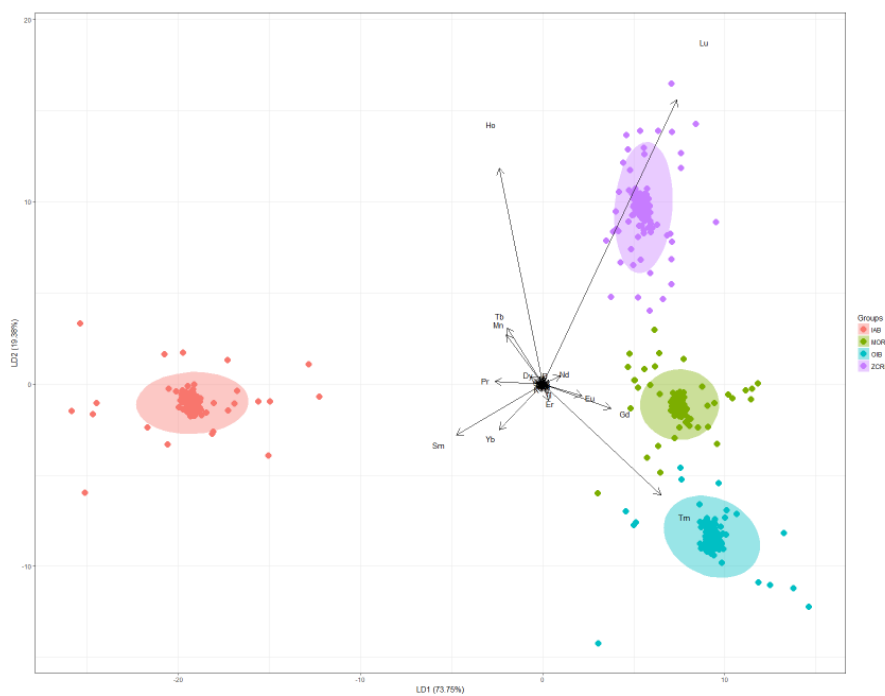
Principal component analysis on rare earth elements (La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu). Plots showing (a) scree plot, (b) loading plot, and (c) scores plot with 95% confidence ellipses around each affinity.



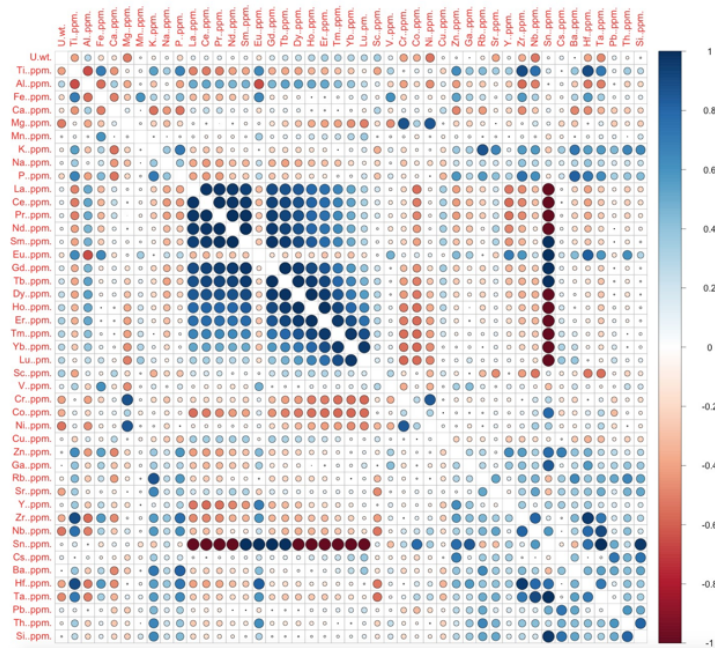
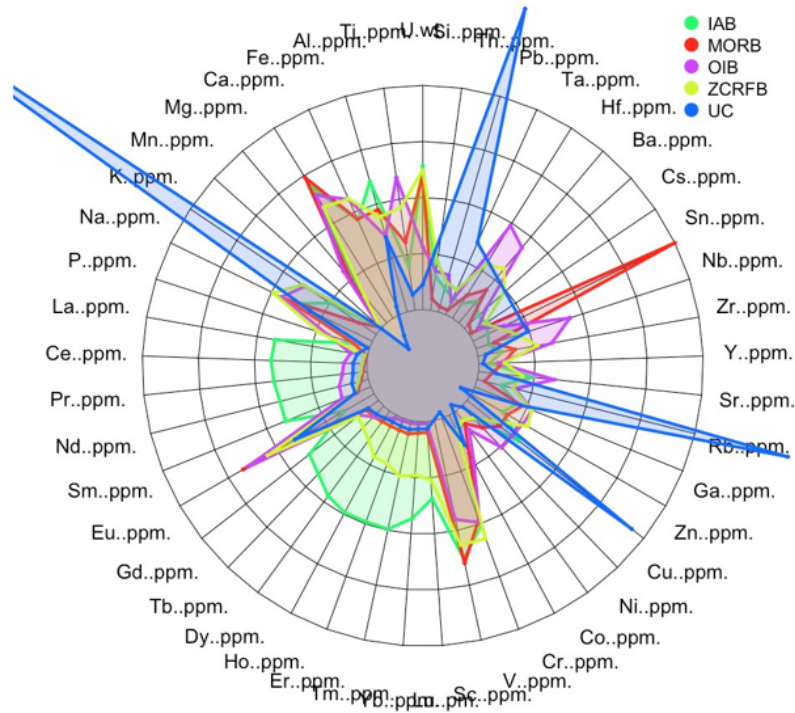
3D presentation of PCA analysis results (REE selected) (ORIGIN-2017)



PCA bi-plot of all known samples for the four affinities using all elements.




LDA bi-plot of samples for the four affinities using all elements.




PLS-DA

Partial Least Squares -Discriminant Analysis (PLS-DA) is a machine learning algorithm that attempts to explain variance through latent variables while simultaneously dividing data into classes. Due to the large amount of missing data in this dataset, imputation was used to fill in the dataset.

 Ricerca Sistema Elettrico	Sigla di identificazione ADPFISS-LP1-123	Rev. 0	Distrib. L	Pag. 35	di 49
--	--	------------------	----------------------	-------------------	-----------------

Acronyms Used

- ALS = Alternative Least Square
- ANOVA = Analysis of variance
- JMP = no acronym, a statistics package offered by Statistical Analysis System (SAS)
- LDA = linear discriminant analysis
- NF = nuclear forensics
- MVA = multivariate analysis
- OPLS-DA = Orthogonal Projection to Latent Structures Discriminant Analysis
- PAST = Paleontological Statistics software package
- PCA = principal component analysis
- PLS-DA = Partial least squares discriminant analysis
- REE = Rare earth element
- SME = Subject matter expert
- SPSS = Statistical Package for the Social Sciences
- UOC = uranium ore concentrate

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	36	49

9 ALLEGATO C: Galaxy Serpent v3 – Summary Report – Phase 2

In Phase 1 of the Galaxy Serpent exercise teams were asked to organize provided surrogate UOC data into a model National Nuclear Forensics Library (NNFL) for the purpose of responding to questions posed by authorities in the course of an investigation involving nuclear or other radioactive (RN) material. The data was derived and repurposed from open-domain geological deep sea core surveys for basaltic core samples. Such data, with the SiO₂ concentration relabeled as uranium and after some statistical manipulation, closely mimics uranium ore concentrate (UOC). By design, the data set was sparse and featured missing elements in order to mirror characteristics of actual UOC data sets and to pose decisions and challenges to participants. Teams responded to prompts regarding the methodologies they used, and their efficacy, in leveraging distinguishing characteristics to discriminate between the four classes of samples represented in the data (IAB = Island Arc Basalt; MORB = Mid-Ocean Ridge Basalt; OIB = Ocean Island Basalt; ZCRFB = Columbia River Flood Basalts). Phase 1 responses were submitted by 23 of the 29 participating teams (79%); this was reported without attribution in the Phase 1 summary report.

In Phase 2, teams were presented with a hypothetical scenario in which three barrels of UOC were found out of regulatory control. Teams were asked to respond to the following queries from authorities about its potential common provenance and consistency to holdings in the model NNFL developed in Phase 1:

1. *Based on a comparison of material characteristics, does the UOC in each of the barrels share a consistent provenance with either of the other two barrels? Describe the methodologies or techniques by which you reached this conclusion.*
2. *Is any of the UOC consistent with the material you recorded in your NNFL during Phase 1? Be specific in reporting whether each individual sample (coded 53023; 8615-OSF-1; 1005) was consistent with your library. Describe the methodologies or techniques by which you reached this conclusion, particularly as it relates to the interrogation of the library created in Phase 1.*
3. *What steps would you, or your government, take after determining the answers to the previous two questions? How would you share the results internally within your own government? What confidence do you have in your reported results, and how did you arrive at this (qualitative or quantitative) confidence level?*

Phase 2 responses were submitted by 22 of the 29 participating teams (76%). Responses to the first two prompts were, in general, more detailed than those for the third prompt. An unattributed summary of the responses from teams is given below in Table 1. **The findings that exercise organizers anticipated is given in the first row in bold: Specifically, a) unknowns “53023” and “8615-OSF-1” are consistent with each other, and b) identified with the IAB affinity in the model NNFL; and c) unknown “1005” was not consistent with the other two unknowns, or d) with any of the four affinities represented in the model NNFL.** All responding teams (22 of 22) came to the same findings on questions a), b) and c). However, some teams (5 of 22) came to a different conclusion on d), finding that unknown “1005” was consistent with one or more of the four affinities in the model NNFL as summarized below. Details on the methods used to reach these findings follow the table.


Table 1: Summary of findings of participant teams

	Barrel #1 – 53023		Barrel #2 - 8615-OSF-1		Barrel #3 – 1005	
	Of consistent origin with other unknowns?	Affinity with NNFL?	Of consistent origin with other unknowns?	Affinity with NNFL?	Of consistent origin with other unknowns?	Affinity with NNFL?
Anticipated	yes	IAB	yes	IAB	no	none
Team #1	yes	IAB	yes	IAB	no	none
Team #2	yes	IAB	yes	IAB	no	none
Team #3	yes	IAB	yes	IAB	no	none
Team #4	yes	IAB	yes	IAB	no	none
Team #5	yes	IAB	yes	IAB	no	none
Team #6	yes	IAB	yes	IAB	no	none
Team #7	yes	IAB	yes	IAB	no	OIB, MORB
Team #8	yes	IAB	yes	IAB	no	none
Team #9	yes	IAB	yes	IAB	no	none
Team #10	yes	IAB	yes	IAB	no	OIB
Team #11	yes	IAB	yes	IAB	no	OIB
Team #12	yes	IAB	yes	IAB	no	none
Team #13	yes	IAB	yes	IAB	no	none
Team #14	yes	IAB	yes	IAB	no	OIB
Team #15	yes	IAB	yes	IAB	no	none
Team #16	yes	IAB	yes	IAB	no	none
Team #17	yes	IAB	yes	IAB	no	none
Team #18	yes	IAB	yes	IAB	no	none
Team #19	yes	IAB	yes	IAB	no	none
Team #20	yes	IAB	yes	IAB	no	none
Team #21	yes	IAB	yes	IAB	no	OIB
Team #22	yes	IAB	yes	IAB	no	none

Discussion of Responses to Prompt 1

Prompt 1 asked: *Based on a comparison of material characteristics, does the UOC in each of the barrels share a consistent provenance with either of the other two barrels? Describe the methodologies or techniques by which you reached this conclusion.*

All teams submitting reports (22 out of 22) found that unknowns 53023 and 8615-OSF-1 shared a common provenance, and that unknown 1005 was not consistent with either unknown 53023 or 8615-OSF-1.

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	38	49

Several teams additionally noted that while unknowns 53023 and 8615-OSF-1 shared a consistent provenance, they were not from the same sample; that is, their provenance was consistent, but not identical.

The details and specificity of the reports varied. Some teams provided summary findings with minimal discussion of how they reached their conclusions, while others provided greater detail as to their methodologies and underlying thought processes.

Missing data

A handful of the reporting teams (~3-5) made some reference to the missing data within the unknowns 8615-OSF-1 and 1005. Two teams discussed this in some detail; one noting three options they considered:

1. Deleting or ignoring the data which they viewed as problematic as it cuts too much of data set)
2. Inserting dummy variables, with they reasoned was not the best option with a large dataset
3. Replacing the missing data with mean or median value of the set. They chose this option for the GSV3 exercise, but noted it can be inaccurate if missing data is not missing completely at random (MCAR) which could be the case here as large blocks of lanthanide data were missing. This affected the confidence of their findings.

Several (2-4) teams appear to have used option 3, as inferred from plots; only two mentioned this directly.

Rare Earth Element (REE) data

Most (19 of 22) teams explicitly noted the use of REE data in their analysis, and others may have done so as well, but not noted this overtly in their report. The main reason some did not utilize REE data seems to be the missing data in unknowns 8615-OSF-1 and 1005. About ten of the teams noted that REE data was the “main discriminant fingerprint,” and provided valuable information as to both origin and history.

Teams found that chondrite normalized REE patterns for unknowns 53023 and 8615-OSF-1 showed significant similarity in pattern, although the REE concentration for unknown 53023 was about a factor of two lower than unknown 8615-OSF-1. Team determined that unknown 1005 was markedly different from the other two unknowns, and thus was not of consistent provenance. Most teams made explicit reference to the Eu anomaly in this regard.

Several teams also cited REE data comparison in their conclusion that while unknowns 53023 and 8615-OSF-1 are of consistent provenance, they are not from the same sample. As evidence, they noted that the normalized values of REE of 8615-OSF-1 sample were approximately twice that of sample 53023, the missing Gd content for unknown 8615-OSF-1, and the variation in the Sr and Ni content between the two.

A couple teams noted that in addition to normalizing to standard chondrite values, normalizing to Tb, a marker of uranium ore deposition, provided further confirmation of their conclusions.

Some examples of the tables and graphs included in reports are provided below.

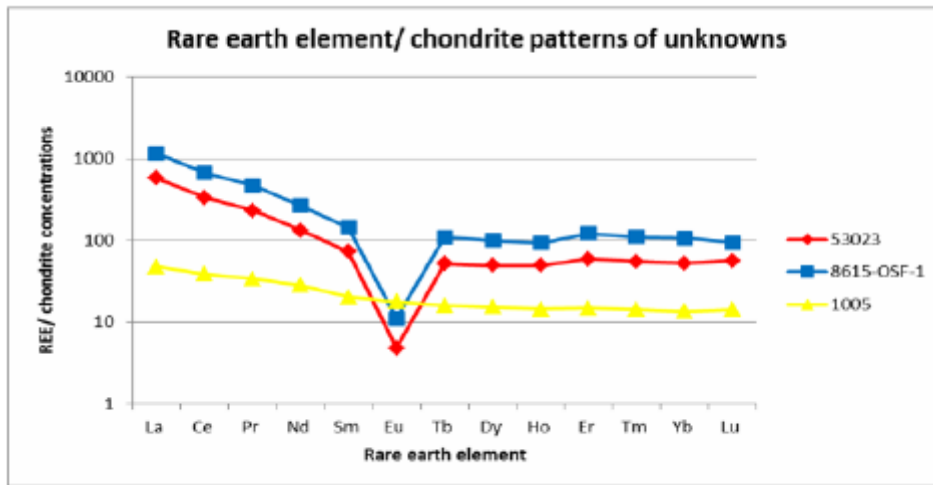


Figure 1: Chondrite normalized REE pattern for the three unknowns

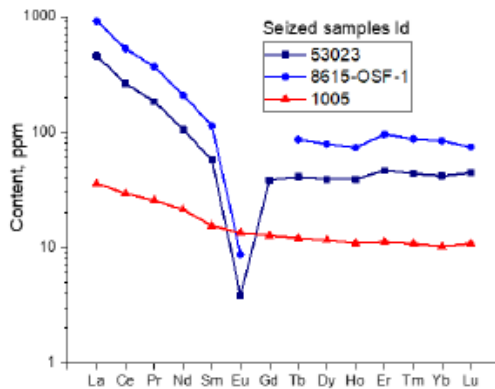


Figure 2: Chondrite normalized REE pattern

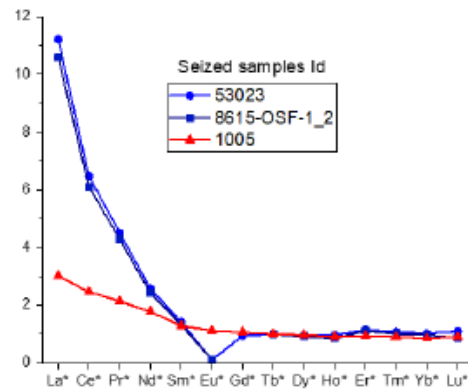


Figure 3: REE pattern normalized to Tb content

Elemental concentrations: 13 of 22 teams explicitly noted the use of non-REE impurity data in their analysis, all in conjunction with REE analysis. Roughly six teams noted challenges associated with the absence of any measurement uncertainties. (The announced exercise guidance was that measurement uncertainties were not provided since variation within class was assumed to be much greater than any analytical uncertainty.)

Many teams noted the uranium concentration of unknowns 53023 and 8615-OSF-1 were similar, while unknown 1005 was slightly lower.

Some examples of the tables and graphs included in reports are provided below.

SAMPLE_ID	53023	8615-OSF-1	1005
Ti (ppm)	0,30	0,30	0,20366
Al (ppm)	17,69	20,68	13,57
Fe (ppm)	7,28	6,53	0,85
Ca (ppm)	12,97	12,64	0,76
Mg (ppm)	6,43	4,64	0,10
Mn (ppm)	0,19	0,16	0,02
Na (ppm)	1,31	1,74	2,17
Y (ppm)	14	13,9	13
Ba (ppm)	87	107	n.d.

n.d. - not determined

AFFINITY	SAMPLE ID	K (ppm)	Nb (ppm)	Pb (ppm)	Th (ppm)
unknown	53023	0,18	0,25	2,00	0,197
unknown	8615-OSF-1	0,23	0,2	1,16	n.d.
unknown	1005	6,26	19	16,5	25,3

n.d. - not determined

Table 2: Tabular comparison of elemental concentrations of three unknowns

AFFINITY	SAMPLE ID	U, wt%	Ti	Al	Fe	Ca	Mg	Mn	K	Na	P	V	Cr	Cu	Ni	Cu	Zn	Ga	Pb	Y	Zr	Nb	Ce	Er	Ba	Hf	Ta	Pb	Th
X01	53023	79.52	0.287	17.89	7.28	12.97	6.43	0.19	0.18	1.31	0.06	309	143	35	44	92	81	13	2.6	14	33	0.25	0.25	358	87	0.94	0.02	2	0.197
X02	8615-OSF-1	79.3	0.296	20.68	6.53	12.64	4.64	0.16	0.23	1.74	0.07				13				3.4	13.9	27	0.2		692	107			1.16	
X03	1005	75.9	0.20366	13.57	0.85	0.76	0.10	0.02	2.17	10.93		246	317	320	325	344	18.35			16.7	13	196.5							

Table 3: Content of elements (REE not included)

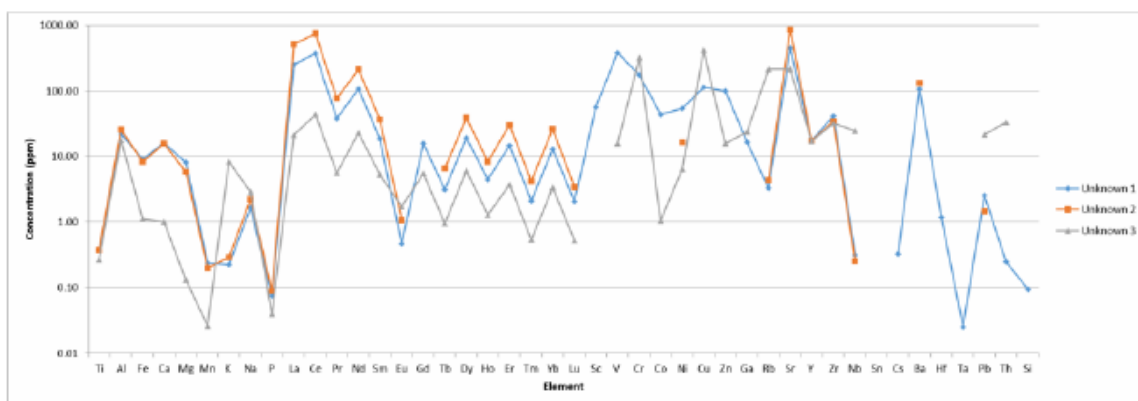


Figure 4: Elemental concentrations of the three unknowns

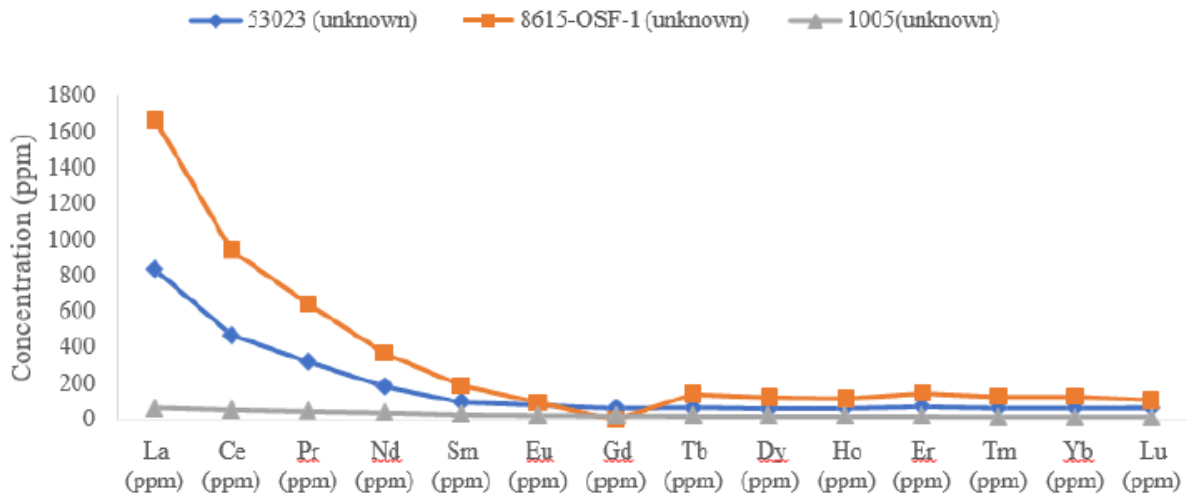


Figure 5: Elemental concentrations of the three unknowns

Relative Standard Deviation (RSD) analysis:


Three teams employed a coefficient of variation analysis as a way to determine if different pairing of two of the three sets of data suggested consistent provenance. The standard deviation of each element represented in the library dataset was calculated. Averages of the four RSDs were then used to compare the relative differences between pairing of unknowns. Scores were assigned, ignoring missing data. For example, one team awarded results within one-RSD three points, results within two-RSDs two points, results within three-RSDs one point, and results greater than three-RSDs no points. These results were then tallied for each pairing of unknowns. The results of such an approach are given below.

	SAMPLE_ID	Elemental Concentrations (ppm)																		Summary Data - Counting occurrences of scores						
		U	Ti	Al	Fe	K	Na	La	Ce	Eu	Tm	Yb	Lu	Ni	Rb	Sr	Y	Nb	Ba	Pb	Th	#3	#2	#1	#0	#no data
Unk 1 Conc	53023	79.5	0.3	17.7	7.3	0.2	1.3	204	300	0.4	1.7	10.4	1.6	44.0	2.6	358	14.0	0.3	87	2.0	0.2					
Unk 2 Conc	8615-OSF-1	79.3	0.3	20.7	6.5	0.2	1.7	408	600	0.8	3.3	20.8	2.7	13.0	3.4	692	13.9	0.2	107	1.2						
Unk 3 Conc	1005	75.9	0.2	13.6	0.9	6.3	2.2	16	33	1.3	0.4	2.5	0.4	4.9	166	167	13.0	19.0		16.5	25.3					
%dif 1&2		0.3	0.4	16	11	24	28	67	67	78	66	67	50	109	27	64	1	22	21	53						
%dif 1&3		5	37	26	158	189	49	171	160	111	121	121	122	160	194	73	7	195		157	197					
%dif 2&3		4	37	42	154	186	22	185	179	42	156	157	149	91	192	122	7	196		174						
Average %RSD*1		2	29	10	13	65	18	38	35	43	25	33	30	79	84	42	29	80	83	96	102					
Average %RSD*2		4	59	20	27	130	37	75	70	85	50	66	60	159	168	85	59	159	165	193	204					
Average %RSD*3		6	88	29	40	195	55	113	105	128	75	99	90	238	252	127	88	239	248	289	307					
Score 1 vs. 2		3	3	2	3	3	2	2	2	2	1	1	2	2	3	2	3	3	3	3		12	12	7	0	14
Score 1 vs. 3		1	2	1	0	1	1	0	0	1	0	0	0	1	1	2	3	1		2	2	3	6	9	20	7
Score 2 vs. 3		1	2	0	0	1	2	0	0	3	0	0	0	2	1	1	3	1		2		3	5	5	17	15

Table 4: Evaluation of similarity or difference of unknowns using relative standard deviations

Discussion of Responses to Prompt 2

Prompt 2 asked: *Is any of the UOC consistent with the material you recorded in your NNFL during Phase 1? Be specific in reporting whether each individual sample (coded 53023; 8615-OSF-1; 1005) was consistent with your library. Describe the methodologies or techniques by which you reached this conclusion, particularly as it relates to the interrogation of the library created in Phase 1.*

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	42	49

The responses to this prompt were the most varied in terms of findings and in reports did not always elucidate the methods used to reach conclusions.

Teams used a variety of methods in interrogating the data for the unknowns, and conducting a comparative analysis with affinities represented in their model NNFL. The results are summarized in Table 5 on the following page. In addition to the methods previously noted, a few teams additionally employed box plots, LDA, Random Forest, and other techniques. Table 5 represents a method-by-method summary of findings. “No” means teams did not find that method as suggesting a given unknown was of consistent provenance with a given affinity; “Possible” means a method found the unknown was possibly consistent with an affinity based on that method alone; and “Yes” indicates a method found the unknown was consistent with an affinity. Numbers in parenthesis indicated the number of teams reaching this finding. Where there are no numbers, all teams reported the finding. Some reports did not provide sufficient detail to make a determination. In general, teams that identified unknown 1005 as being consistent with an affinity in the model NNFL appear to have utilized one comparative method. In a handful of cases, it was difficult to determine whether a team was reporting a provenance match as a confident finding, or a suggested possible finding worthy of further examination. In such cases “best possible inferences” were made.

While all teams determined that unknown 53023 and 8615-OSF-1 were consistent with the IAB affinity with a high degree of confidence, the majority noted that the finding was less strong for unknown 8615-OSF-1.

Only three teams made mention that unknown 8615-OSF-1 exactly matched a specific IAB sample identification in the model library. Those that compared these directly reported that the elemental concentrations were consistent, with the exception of Sr, but not identical.

Unknown 53023				
Affinity	REE concentrations	Non-REE concentrations	PCA on REE	PCA on non-REEs
IAB	Yes	Yes	Yes	Yes
OIB	No	No	No	No/Possible (1)
MORB	No	No	No	No
ZCRFB	No	No	No	No/Possible (1)
Unknown 8615-OSF-1				
Affinity	REE concentrations	Non-REE concentrations	PCA on REE	PCA on non-REEs
IAB	Yes	Yes	Yes	Yes
OIB	No	No	No	No/Possible (1)
MORB	No	No	No	No
ZCRFB	No	No	No	No/Possible (1)
Unknown 1005				
Affinity	REE concentrations	Non-REE concentrations	PCA on REE	PCA on non-REEs
IAB	No	No/Possible (1)	No/Possible (1)	No
OIB	No/Possible (3) /Yes (4)	No/Possible (2) /Yes (1)	No/Possible (2) /Yes (1)	No/Possible (1)
MORB	No/Possible (2) /Yes (1)	No/Possible (1)	No/Possible (1)	No
ZCRFB	No	No	No/Possible (2)	No/Possible (2)

Table 5: Results of different methodologies applied by participant teams in Phase 3

Rare Earth Element (REE) data

Rare earth elemental comparison appears to have been used by the majority, if not all, of reporting teams. The Eu anomaly and slope across all elements exhibited by unknowns 53023 and 8615-OSF-1 was also present in affinity IAB. None of the other affinities displayed the Eu anomaly. Teams cited this as the primary evidence in concluding that 53023 and 8615-OSF-1 share a common provenance with affinity IAB. Many noted that this conclusion was stronger with 53023 than with 8615-OSF-1 due to slight differences in Y concentrations, and the missing Gd signature. Many found the third unknown, 1005, shared no similarity in terms of specific features or slope to any other affinity, leading most teams to conclude it was extremely unlikely to share a common provenance with any known affinity in the model NNFL. However, some teams noted that the REE pattern of 1005 bore some similarity to OIB for

the lighter REEs, and to MORB for the heavier REEs, leading them to include these as a potential common provenance.

Sample plots for this section are shown below.

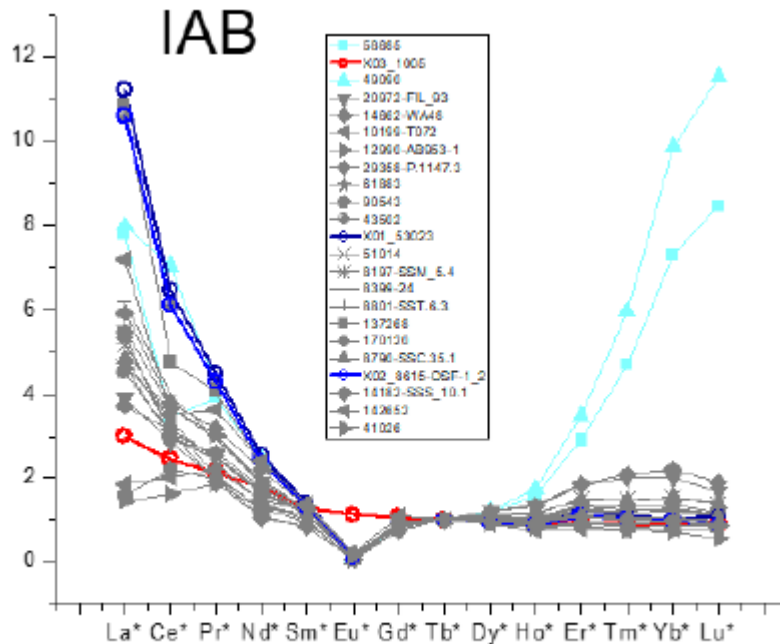


Figure 6: Chondrite normalized REE data

Non-REE concentration comparison

Results here were similar to REE concentration comparative analysis, but the reported findings were more consistent across teams. Teams found that 53023 and 8615-OSF-1 were of consistent provenance as affinity IAB, with stronger confidence for 53023 than with 8615-OSF-1 due to slight differences in K, Ni, Nb, Sr and Zr concentrations. In general, unknown 1005 was not found to be consistent with any affinity by this method.

AFFINITY	SAMPLE_ID	U_wt%	Ti	Al	Fe	Ca	Mg	Mn	K	Na	P	V	Cr	Co	Ni	Cu	Zn	Ga	Rb	Y	Zr	Nb	Cm	Sr	Ba	Hf	Ta	Pb	Th
X01	53023	79.52	0.297	17.69	7.28	12.97	6.43	0.19	0.16	1.51	0.06	309	143	35	44	92	31	13	2.6	14	30	0.25	0.26	359	67	0.94	0.02	2	0.197
X02	8615-OSF-1	79.3	0.296	20.68	6.53	12.54	4.94	0.16	0.23	1.74	0.07				13				3.4	13.9	27	0.2		692	107			1.18	
X03	1005	75.9	0.20365	12.67	0.65	0.76	0.4	0.32	6.26	2.47	0.03	12	246	6.8	4.9	305	12		10.35	159	13	26.8	15		166.5			16.6	25.4

Table 6: Elemental content (REE not included)

AFFINITY	Sample_ID	U_wt_per	Ti_ppm	Al_ppm	Ca_ppm	Mg_ppm	Mn_ppm	K_ppm	Na_ppm	P_ppm	Ni_ppm	Rb_ppm	Sr_ppm	Y_ppm	Zr_ppm	Nb_ppm	Ba_ppm	Pb_ppm	CaSr	BaSr
IAB	8615-OSF-1	79.3	0.296	20.68	12.54	4.64	0.16	0.23	1.74	0.07	13	3.4	102	13.9	27	0.2	107	1.15	6.53E-02	0.657
X02	NFS-8615-OSF-1	79.3	0.206	20.68	12.54	4.64	0.16	0.23	1.74	0.07	13	3.4	602	13.9	27	0.2	107	1.15	1.83E-02	0.165

Table 7: Elemental comparison of an IAB sample to 8615-OSF-1

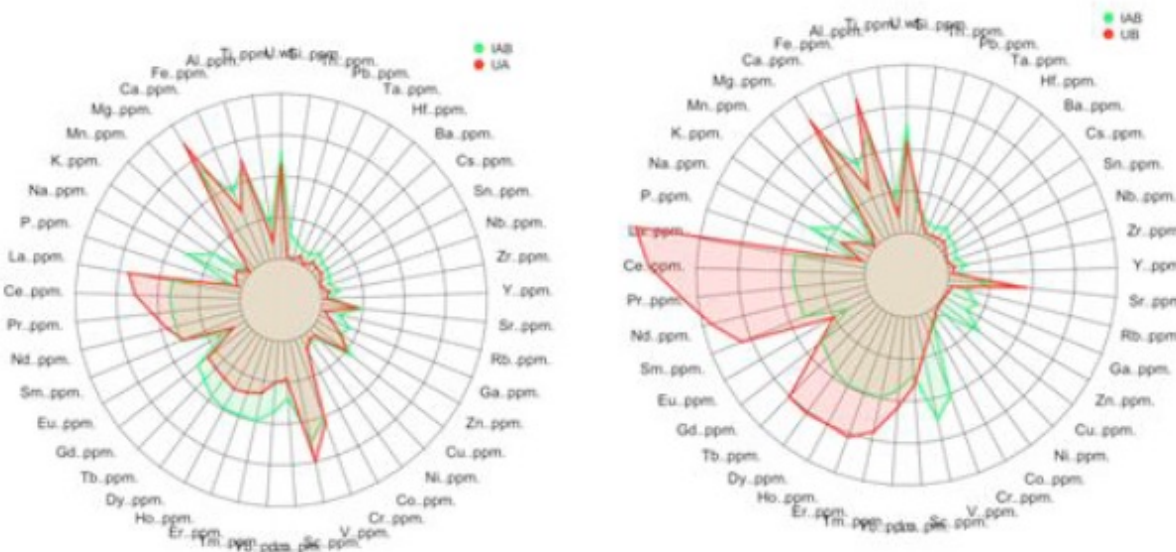


Figure 7: Radar Plot Assessments of Samples 53023 (left) and 8615-OSF-1 (right)

PCA - REE

Unknowns 53023 and 8615-OSF-1 were found to be nearest the IAB affinity by this method. Both fell within the known IAB sample points, while teams found that 53023 was consistently closer to the center of the IAB cluster than unknown 8615-OSF-1, giving relatively higher confidence. Most teams found that unknown 1005 did not lie within the most distant scattered points for any affinity in the model NNFL, however about five teams determined that unknown 1005 lay within the OIB, MORB or ZCRFB clusters, suggesting that, by this method, it was possible that 1005 was consistent with those affinities. One team suggested this consistency was sufficient to report a consistent provenance with affinity OIB.

PCA on non-REEs

Teams using this method generally found that unknowns 53023 and 8615-OSF-1 were of consistent provenance as the IAB affinity, and unknown 1005 did was not consistent with any of the four affinities. A few groups found possible correlations of unknowns with certain affinities, as summarized in the table above.

Sample PCA plots (and one PLS-DA) are included below.

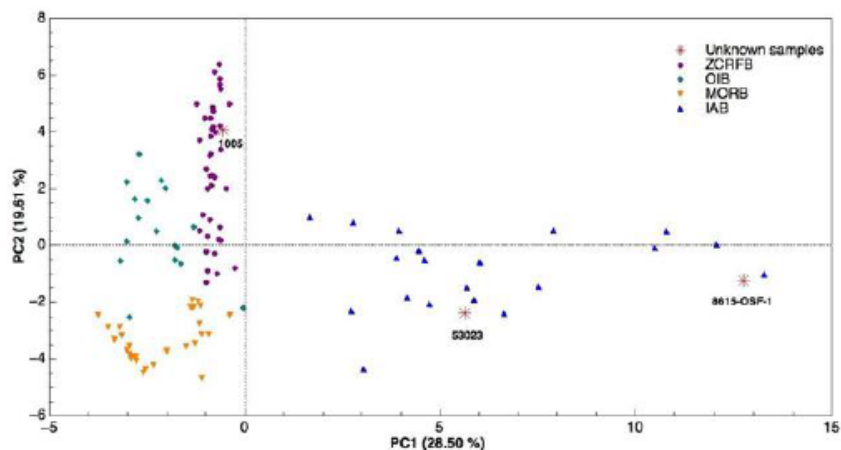


Figure 8: PCA Analysis of unknowns and model NNFL

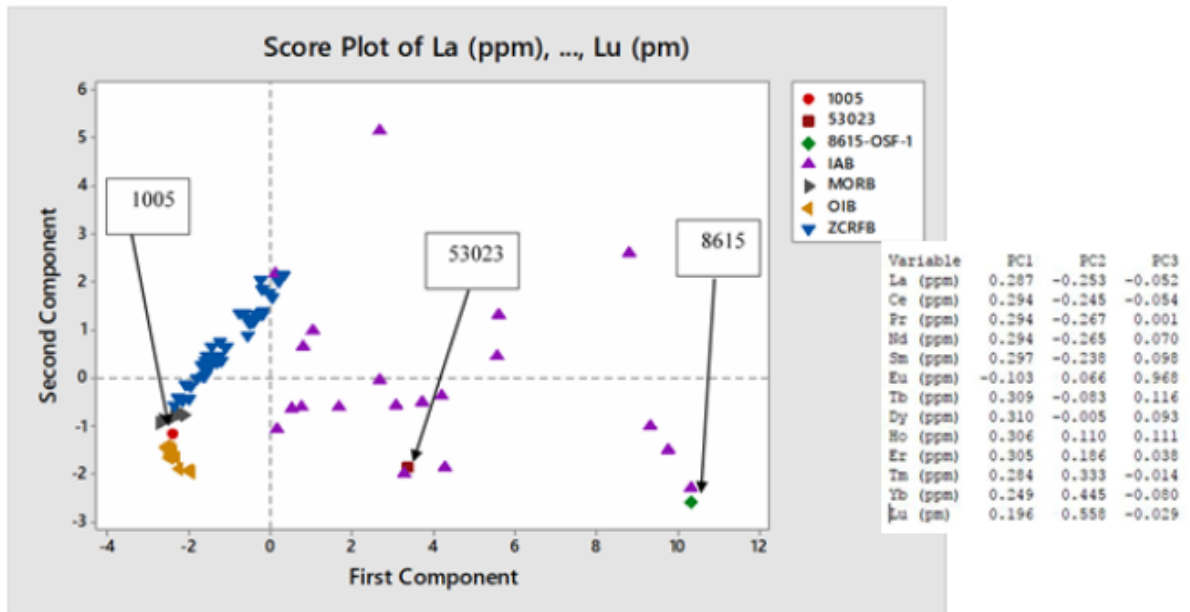


Figure 9: PCA for REE without Gd

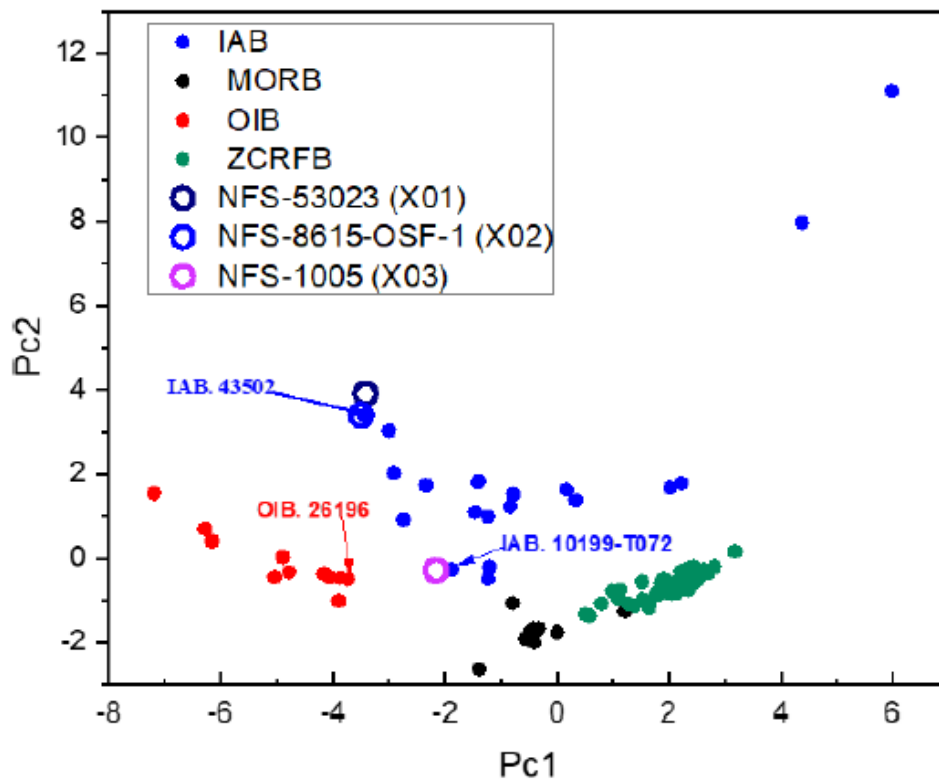


Figure 10: PCA for REEs

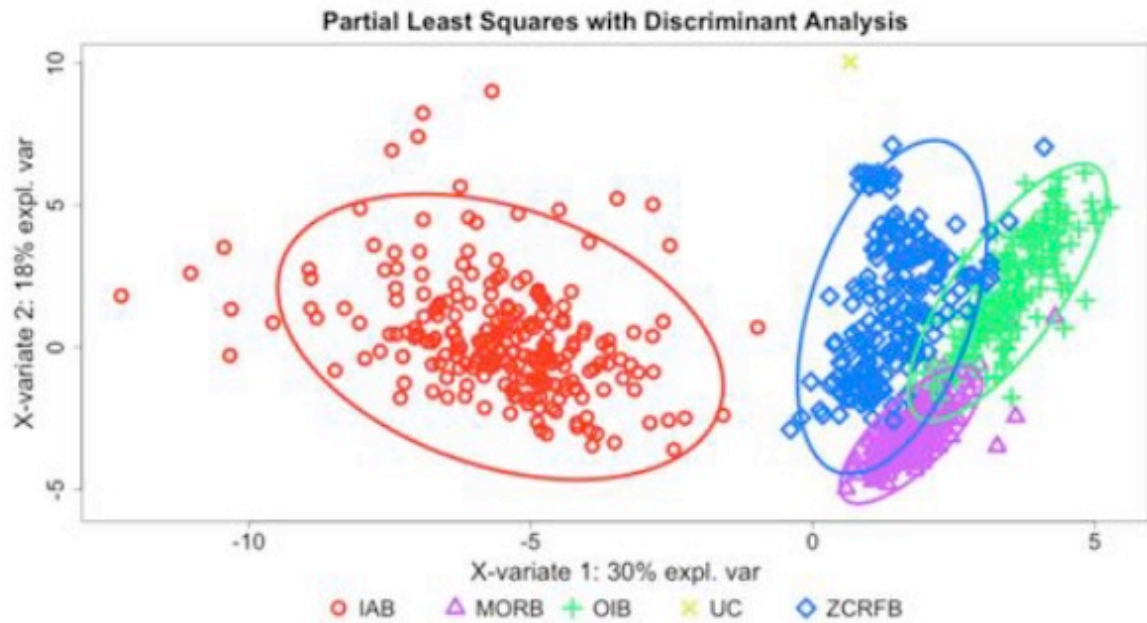


Figure 11: Partial Least Squares with Discriminant Analysis of Sample 1005

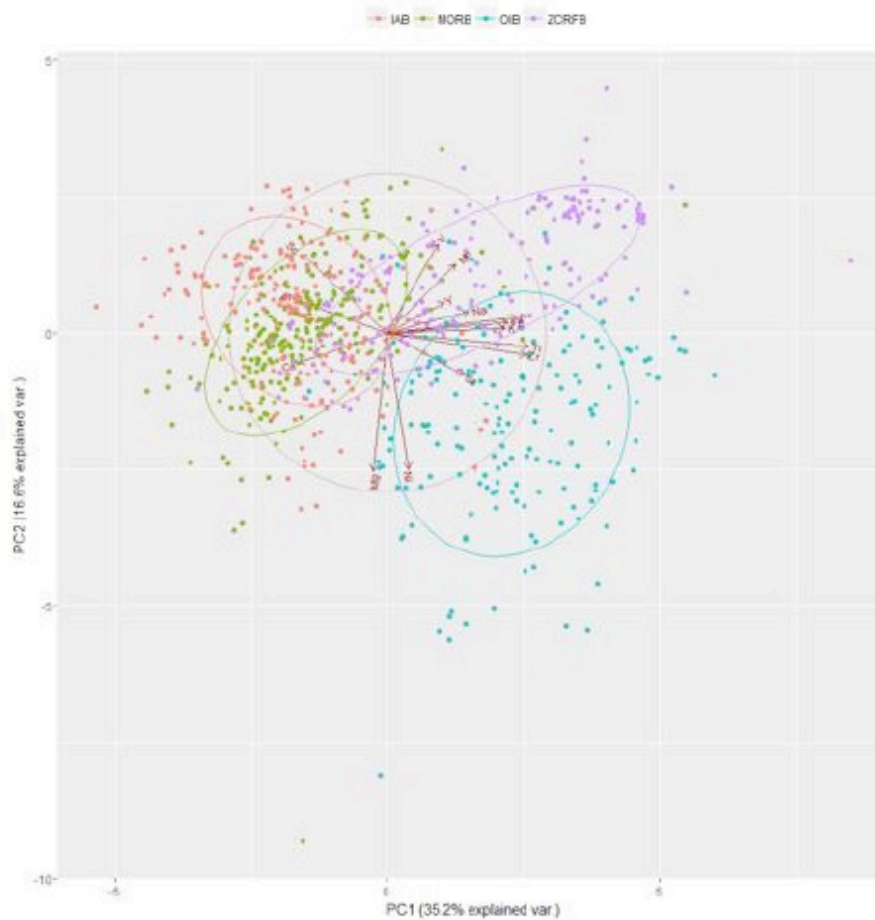




Figure 12: PCA bi-plot of all samples in the four affinities using the 15-element subset.

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	48	49

Discussion of Responses to Prompt 3

Prompt 3 asked: *What steps would you, or your government, take after determining the answers to the previous two questions? How would you share the results internally within your own government? What confidence do you have in your reported results, and how did you arrive at this (qualitative or quantitative) confidence level?*

Comments in this section were brief and varied. Most teams reiterated their findings and noted they would report them to the appropriate governmental agency making the request. Some teams noted they would suggest the governmental authority contact known sites involved in the IAB UOC mining and milling processes to check inventory. Others reported they would recommend additional chemical analysis on the unknowns. A number of teams noted they do not comment on governmental processes. Several noted that the purpose of the lab is to answer investigative questions, and not advise on any additional steps.

	Ricerca Sistema Elettrico	Sigla di identificazione	Rev.	Distrib.	Pag.	di
		ADPFISS-LP1-123	0	L	49	49

10 Bibliografia

- [1] "IAEA NSS2 Nuclear Forensic Support Technical Guidance Reference Manual".
- [2] "Proposed Framework for National Nuclear Forensic Libraries and International Directories", J. F. Wacker (Pacific Northwest National Laboratory), U.S. Department of Energy, Michael Curry (U. S. Department of State) – 08 June 2011.
- [3] "IAEA NSSXX Development of a National Nuclear Forensics Library", DRAFT, February 2013.
- [4] "Radionuclide and Radiation Protection Data Handbook 2002". Radiation Protection Dosimetry vol 98 No. 1, 2002 - Nuclear Technology Publishing.
- [5] "Introduzione all'analisi degli errori", John Taylor, Zanichelli 1996.
- [6] "Elementi di probabilità e statistica", P. Galeotti, Levrotto & Bella 1983.
- [7] "Metodi di Analisi Statistica Multivariata", D. F. Morrison, Casa Editrice Ambrosiana, 1976
- [8] "Complementi di Matematica e Statistica", Lubisco A (Università degli Studi di Bologna, 2009-2010).