

Trabajo Final de Grado

Grado en Ingeniería Química

Establecimiento de la procedencia de muestras arqueológicas de obsidiana de Oriente Próximo mediante el programa estadístico Source Seeker 2.1

MEMORIA

Autor: Adrián López Fernández
Director: Francisco Javier Giménez Izquierdo;
Luis Solano Albajes
Convocatoria: Enero 2017



Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona



Resumen

El presente trabajo tiene como objetivo principal continuar con el desarrollo del programa informático Source Seeker destinado a la realización de análisis de procedencia basados en el estudio geoquímico de la obsidiana. Se centra en implementar algunas de las mejoras sugeridas por sus predecesores y aquellas necesarias para la mejora de su funcionamiento.

Los estudios de procedencia se basan en la existencia de sustancias químicas o características mineralógicas que dependen de la ubicación geográfica del material. Forman parte de los campos más importantes donde se comprueba la utilidad de la química en ramas más humanistas, como la arqueología o la historia. Los resultados obtenidos permiten delimitar el tamaño de los territorios, las interacciones entre las diferentes culturas o civilizaciones y las posibles rutas comerciales existentes en la antigüedad.

Por lo tanto, este trabajo pretende centrarse en el estudio explicado anteriormente, y se dedicará parte de su contenido a realizar una mejora del programa informático, así como a facilitar su uso para que se puedan realizar análisis de procedencia sin poseer amplios conocimientos informáticos. Del mismo modo, se pretende analizar material arqueológico de Oriente Próximo y Medio con el fin de comprobar la aplicabilidad del programa desarrollado.

Sumario

RESUMEN	2
SUMARIO	4
1. GLOSARIO	7
2. ORIGEN DEL PROYECTO Y MOTIVACIONES	9
3. INTRODUCCIÓN	10
3.1. Objetivos del proyecto	11
3.2. Alcance del proyecto.....	11
4. DESARROLLO DEL TRABAJO	13
4.1. Análisis previo	13
4.1.1. Estudio del análisis de procedencia	13
4.1.2. Vías de continuación propuestas	13
4.1.3. Estudio inicial	14
4.1.4. Conclusiones del estudio inicial	15
4.2. Ampliación de las fuentes	16
4.3. Mejora del programa.....	18
4.3.1. Random Forest	18
4.3.2. Selección de variables	19
4.4. Ejecutable	21
4.5. Instalador	24
5. ANÁLISIS DE LOS ARTÍCULOS	26
5.1. Cálculos previos.....	26
5.2. Localización de los yacimientos.....	27
5.2.1. Artículo 1.....	27
5.2.2. Artículo 2.....	30
5.2.3. Artículo 3.....	33
5.2.4. Artículo 4.....	36
5.2.5. Artículo 5.....	40
5.2.6. Artículo 6.....	45
5.2.7. Artículo 7.....	48
5.2.8. Artículo 8.....	50
5.2.9. Artículo 9.....	54
5.2.10. Artículo 10.....	57
5.3. Artículos nuevos	60

5.3.1. Tell el-Iswid	60
5.3.2. Tell-el-fara	63
6. VÍAS DE CONTINUACIÓN	68
6.1. Soporte para otras extensiones	68
6.2. Ampliación de las fuentes existentes	68
6.3. Registro de nuevas fuentes.....	69
6.4. Elipsoide de error	69
6.5. Exportar resultados	70
6.6. Creación de una página web.....	70
7. COSTES	71
8. PLANIFICACIÓN	72
9. IMPACTO AMBIENTAL	74
CONCLUSIONES	75
AGRADECIMIENTOS	76
BIBLIOGRAFÍA	77
Referencias bibliográficas.....	77
ANEXO I	79

1. Glosario

Análisis por activación neutrónica (NAA): Consiste en irradiar una muestra para determinar la composición de sus elementos con neutrones térmicos, por lo general en un reactor nuclear de investigación.

Anatolia: Península ocupada actualmente por la parte asiática de Turquía. Limita al norte con el mar Negro, al sur con el mar Mediterráneo, al oeste con el mar Egeo y al este con el resto del continente asiático.

Bug (o error de software): Error o fallo en un programa de ordenador o sistema de software que desencadena un resultado indeseado.

Cantera: Explotación minera, generalmente a cielo abierto, en la que se obtienen rocas industriales, ornamentales o áridos.

Capadocia: Región histórica de Anatolia Central, en Turquía. Se caracteriza por tener una formación geológica única en el mundo y por su patrimonio histórico y cultural.

Fluorescencia de rayos X (XRF): Consiste en emisión de rayos X secundarios característicos de un material que ha sido excitado al ser incidido con rayos X de alta energía o rayos gamma.

Fuente: Lugar natural de donde sale o emana algo. En el caso del presente trabajo se refiere al lugar de donde procede la obsidiana.

Minería de datos: Campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

Módulo (informática): Porción de un programa de ordenador.

Obsidiana: Roca ígnea volcánica con alto contenido en sílice de composición parecida al granito. Es un material duro y fácilmente laminable.

Python: Lenguaje de programación interpretado, usa tipado dinámico y es multiplataforma. Soporta orientación a objetos, programación imperativa y programación funcional.

Scripts (o archivo de procesamiento por lotes): Programas, usualmente pequeños o simples, para realizar tareas muy específicas, escritas en códigos de programación, generalmente almacenadas en un archivo de texto.

Shell (o intérprete de comandos): Programa informático que provee una interfaz de usuario para acceder a los servicios del sistema operativo.

Valor perdido: Valor que no fue recogido durante el proceso de medición.

Yacimiento (arqueológico): Concentración de restos arqueológicos en el cual podemos encontrar restos de actividad humana. Está constituido por la presencia de artefactos, elementos estructurales, suelos de ocupación y otra serie de anomalías.

2. Origen del proyecto y motivaciones

El planteamiento inicial del proyecto surgió de la propuesta del profesor Francisco Javier Giménez realizada en la asignatura Desenvolupament Tecnològic i Científic a l'Antiguitat. Egipte i Orient Pròxim, impartida en la Escuela Técnica Superior de Ingeniería Industrial de Barcelona, ETSEIB. Este proyecto es la continuación de dos proyectos anteriores, realizados por Mauricio Andrés Alva Howes [17] y Joel Moriana Becerra [1].

Ambos proyectos se basaron en la creación y mejora de un programa informático cuyo objetivo principal es realizar estudios de procedencia basados en la composición química de una serie de muestras de obsidiana. Gracias a que la obsidiana, o vidrio volcánico, presenta una composición química diferente según su fuente de procedencia, se puede determinar el lugar de procedencia de las muestras arqueológicas estudiando la composición de dichas muestras y utilizando una base de datos de composición de obsidiana de diferentes volcanes.

Este proyecto está centrado en la continuación del desarrollo del programa, así como en la verificación de su funcionamiento en referencia a la identificación de las fuentes de procedencia. La principal motivación fue la de realizar un análisis exhaustivo de un conjunto de muestras para verificar el buen funcionamiento del programa y las posibles deficiencias que este podría tener. Para la realización del análisis se estudió una batería de muestras de obsidiana ya publicadas, una serie de 10 artículos que contenían información y material arqueológico analizado proveniente de Oriente Próximo. Dichos artículos fueron proporcionados por el director del proyecto. Se implementaron una serie de mejoras en el programa con tal de realizar el análisis de procedencia de manera más exhaustiva y cómoda.

Por último, en el presente proyecto se realiza un análisis de dos artículos cuyas muestras encontradas en Egipto son de procedencia geológica desconocida, con tal de poner en práctica el programa y aclarar el origen de dichas muestras.

3. Introducción

El análisis químico de diferentes materiales antiguos puede proporcionar información de su procedencia. En estos casos, la comparación entre la composición química del material de una cantera o una mina y el material encontrado en un yacimiento arqueológico permite la identificación de las fuentes más probables de las que proviene el material arqueológico[11].

Estos estudios, los llamados estudios de procedencia, forman parte de los campos más importantes donde se comprueba la utilidad de la química en ramas más humanistas, como la arqueología o la historia. Los resultados obtenidos permiten delimitar el tamaño de los territorios, las interacciones entre las diferentes culturas o civilizaciones y las posibles rutas comerciales existentes en la antigüedad.

Los estudios de procedencia se basan en la existencia de sustancias químicas o características mineralógicas que dependen de la ubicación geográfica del material, es decir, difieren de un sitio geológico a otro. Estas características podrían ser la concentración mayor o menor de determinados componentes o la proporción de los diferentes minerales que componen la roca.

La obsidiana o vidrio volcánico, es una roca volcánica con un contenido relativamente alto de silicio que tiene una serie de características ventajosas para la realización de los estudios de procedencia. Por un lado, fue muy utilizada en la antigüedad y existen muchas muestras arqueológicas[12]. Debido a su alta dureza y facilidad para ser laminada se utilizó como herramienta de corte en el pasado hasta su utilización en bisturís quirúrgicos en la actualidad. También fue utilizada como piedra preciosa formando parte de ornamentos muy valorados o decorativos en gran parte del Mediterráneo, así como en el resto del mundo, desde Asia hasta América.

Por otro lado, existe un número limitado de minas de obsidiana trabajadas en la antigüedad, de modo que a priori no existe una gran dispersión de composición química en las muestras, favoreciendo la delimitación de los valores químicos que definen cada mina. El alto número de artefactos arqueológicos de obsidiana encontrados en los yacimientos y el reducido número de minas explotadas en el pasado hacen de esta una perfecta muestra de estudio[11].

3.1. Objetivos del proyecto

El objetivo principal de este proyecto consiste en verificar el funcionamiento del programa con el fin de obtener una buena herramienta informática para la realización de estudios de análisis de procedencia. El análisis se dividirá en dos fases:

1. Estudio de muestras ya identificadas.
2. Caracterización de muestras de origen desconocido.

El primer objetivo está dividido en dos partes. La primera parte consiste en identificar las posibles mejoras a implementar en el programa informático con el fin de poder realizar con más precisión y comodidad dichos estudios de análisis de procedencia.

La segunda parte radica en testear los resultados proporcionados por el programa mediante una batería de muestras extraídas de diferentes artículos científicos. Utilizando la base de datos de la que disponemos, se analizarán dichas muestras para verificar sus resultados y compararlos con las conclusiones extraídas de cada artículo en cuestión.

Por último, gracias al uso de esta herramienta informática se procederá al análisis de dos muestras de las cuales se desconoce su origen con el fin de esclarecer su fuente de procedencia. Para ello se utilizarán las mejoras implementadas en el programa para lograr su correcta identificación.

3.2. Alcance del proyecto

La intención de este proyecto, del mismo modo que sus predecesores, es que se pueda continuar con su desarrollo en el futuro, mejorándolo progresivamente.

En este trabajo se lleva a cabo la mejora y perfección del programa desarrollado anteriormente. Con el fin de poder realizar dicha tarea el presente proyecto abarcará los aspectos que se muestran a continuación:

- Mejora en el proceso de selección de muestras en los diferentes tipos de análisis.
- Implementación de una tabla de resultados en el análisis de procedencia de las muestras.
- Desarrollo de un ejecutable válido y sencillo para una fácil distribución y utilización.
- Verificación del adecuado funcionamiento del programa.
- Realización del estudio sobre las muestras facilitadas.
- Comparativa de los resultados extraídos del programa con los artículos científicos.
- Identificación de las muestras de fuentes desconocidas.

- Descripción de las posibles vías de continuación.

4. Desarrollo del trabajo

4.1. Análisis previo

Una vez definido el proyecto, el primer punto a tratar consistió en realizar una lectura detallada de la memoria del proyecto anterior[1] así como la familiarización con el funcionamiento del programa Source Seeker. Se estudiaron las diferentes vías de continuación propuestas y se realizó un primer análisis de procedencia de obsidiana con dicho programa con tal de poder observar qué propuestas de mejora era necesario implementar.

4.1.1. Estudio del análisis de procedencia

El programa Source Seeker se centra en el análisis de material arqueológico de Oriente Próximo y Medio basándose en una base de datos que contiene las diferentes fuentes de procedencia de la obsidiana. Dada la imposibilidad de realizar el muestreo y posterior análisis químico de cada muestra de obsidiana, las muestras de estudio son extraídas de una serie de artículos proporcionados por el director del proyecto.

4.1.2. Vías de continuación propuestas

Existen diferentes propuestas de mejora por parte del anterior proyectista, las cuales pueden clasificarse en nuevas funciones del programa y propuestas de tratamiento datos.

Nuevas funciones

- *Soporte para otras extensiones*: Soporte para la nueva extensión xlsx de Excel y compatibilización de la importación de ficheros de extensión csv, *comma-separated values*, para poder generar datos propios sin necesidad de utilizar Excel.
- *Autocompletado y validación en el graficado*: Permite autocompletar cualquiera de los compuestos introducidos en el gráfico y la validación automática de una función.
- *Barra de progreso*: Permite mostrar el estado de avance de una tarea o proceso, evitando que el programa entre en modo “*No responde*” cuando este realiza algoritmos con un gran número de variables.
- *Selección de variables*: Permite la disminución de la dimensión de un conjunto de datos proporcionando una representación de los datos de manera más simple compacta.

Propuestas de tratamiento de datos

- *Elipsoide de error*: Permite la observación de un tercer compuesto o función introduciendo el cálculo del elipsoide de error en el espacio tridimensional.
- *Mejora del Random Forest*: Incluye el cálculo del número de veces que una muestra determinada se ha predicho en una fuente, permite decidir sobre la certeza de una predicción.
- *K-nearest neighbors*: Permite clasificar una muestra en un conjunto de datos, según la distancia mínima entre la muestra y el conjunto estudiado.
- *Ensembles*: Incluye modelos de clasificación para mejorar el rendimiento de una predicción.
- *Otros algoritmos*: Incluir diferentes algoritmos de clasificación que permitan mejorar cualquier clasificación de datos.

Otras propuestas

El director del proyecto también propuso algunas mejoras a realizar en el programa, las cuales fueron:

- *Realizar un ejecutable*: Se propuso crear un ejecutable de manera que fuera rápido y sencillo utilizar el programa sin necesidad de acceder a *Python*.
- *Desarrollar un instalador*: La creación de un instalador permite la publicación y utilización del programa por parte de cualquier usuario.

4.1.3. Estudio inicial

Para decidir que propuestas era más necesario implementar y cuáles de ellas eran las más viables, se realizó un estudio preliminar del programa. Dicho estudio consistía en realizar un análisis de procedencia de diferentes muestras de obsidiana extraídas de los artículos proporcionados por el director del proyecto. Para empezar con el estudio se analizaron los cinco primeros artículos, los artículos [2], [3], [4], [5], y [6].

Un primer análisis puso de manifiesto la escasez de fuentes de obsidiana en la base de datos del programa. Source Seeker solo posee una base de datos con las fuentes Nemrut Dağ, Göllü Dağ y Bingöl. Dichos artículos contenían muestras las cuales podían provenir de otros volcanes desconocidos por el programa.

La escasez de fuentes de procedencia de obsidiana conlleva que el programa no pueda asignar correctamente las muestras a sus posibles fuentes de origen. No obstante, se disponía de una gran variedad de artículos y datos de diferentes fuentes para incorporar al programa que facilitarían la correcta asignación de las distintas muestras.

Otro aspecto importante fue encontrar diferentes programas para poder realizar el ejecutable requerido. Existen diferentes programas para Python que convierten los scripts de Python en programas ejecutables en Windows. Dichos ejecutables son capaces de funcionar sin requerir una instalación previa de Python.

4.1.4. Conclusiones del estudio inicial

Una vez concluido el estudio preliminar del programa, se descartaron algunas de las propuestas descritas anteriormente y se fijaron otras nuevas.

Nuevas funciones

Con tal de poder realizar análisis más individuales de cada muestra se decidió implementar un método de selección de variables, tanto en el Random Forest como en el graficado. De este modo se garantiza una mejor decisión de asignación de muestras y la posibilidad de realizar análisis en mayor profundidad.

Otra función interesante era añadir una barra de progreso en los cálculos, sin embargo, en el estudio inicial se observó una mejora introducida por el anterior proyectista que evitaba que el programa entrara en modo “*No responde*” cuando los cálculos eran excesivamente largos. Mediante la utilización de una pantalla de espera el programa funcionaba correctamente, de tal forma que se descartó añadir ninguna otra función.

Propuestas de tratamiento de datos

En primer lugar, se descartó implementar cualquier otro tipo de método estadístico y nuevos algoritmos de clasificación de datos, ya que el programa dispone de suficientes herramientas muy completas y cuyas decisiones son lo suficientemente precisas.

Sin embargo, se decidió mejorar el Random Forest de manera que este proporcionara una tabla resumen incluyendo el cálculo del número de veces que una muestra se ha predicho en una fuente. De esta forma, y complementándolo con la introducción del método de selección de variables nombrado anteriormente, se podría realizar un análisis individual de una muestra proporcionando una predicción de mayor rigor sobre la fuente de origen.

Otras propuestas

La creación de un ejecutable y un instalador adecuado para el programa es una de las principales mejoras que se requerían, con ellas se permitiría compartir el programa, instalarlo y ejecutarlo sin necesidad de conocer el funcionamiento y los comandos del lenguaje Python. Por este motivo se decidió llevar a cabo su desarrollo.

Finalmente, la necesidad de obtener nuevas fuentes se hizo patente en el estudio inicial al observar que gran parte de los artículos contenían nuevas fuentes de origen interesantes para el estudio de la procedencia de muestras de obsidiana.

Por lo tanto, se decidió que la lista final de mejoras a implementar fuera la siguiente:

- Añadir nuevas fuentes.
- Mejorar el Random Forest añadiendo una tabla resumen.
- Funciones Seleccionar y Deseleccionar muestras.
- Realizar un ejecutable.
- Desarrollar un instalador.

4.2. Ampliación de las fuentes

Una de las primeras mejoras a realizar fue la ampliación de las fuentes del Excel utilizado para el análisis de procedencia de muestras de obsidiana. En total se añadieron tres nuevas fuentes:

- Ziyaret: Situado en la zona del Lago Van (lago que se encuentra al este de la península de Anatolia, en Turquía, cerca de la frontera con Irán).
- Pasinler: Situado en la zona del Lago Van.
- Nenezi Dağ: Situado en la Capadocia, cerca de la fuente Göllü Dağ.

Tanto los datos de la primera fuente, Ziyaret, como de la segunda, Pasinler, fueron extraídos del tercer artículo analizado [4], el cual contenía información de la composición química de dichas fuentes junto con la composición química de las muestras arqueológicas de obsidiana.

Para la tercera y última fuente añadida se utilizaron tres nuevos artículos como fuente de datos de este nuevo volcán, Nenezi Dağ. De los artículos [13], [18] y [20] se recopilaban nueve muestras diferentes de esta fuente. A causa de disponer diferentes artículos con datos geológicos de la obsidiana de la fuente Nenezi Dağ se decidió realizar un estudio de dichos datos ante la posibilidad de que esta fuente presentara algún subgrupo como es el caso del volcán Nemrut Dağ.

Antes de aplicar el algoritmo de agrupamiento *k-means*, se realizó un estudio previo de los datos de dicha fuente. Para ello se utilizó la función de *Mejores combinaciones* (véase Figura 4.1) del programa y se graficaron los resultados obtenidos.

Compuesto	Compuesto	Indice
Nb [ppm]	Y [ppm]	261.277052202
Sr [ppm]	Y [ppm]	246.108083075
Rb [ppm]	Zr [ppm]	169.798541954
Sr [ppm]	Zr [ppm]	155.455995932
Y [ppm]	Zr [ppm]	155.240394534

Fig. 4.1. Resultados algoritmo mejores combinaciones

Las combinaciones realizadas no mostraron que la nueva fuente Nenezi Dağ presentará ningún subgrupo (véase Figura 4.2).

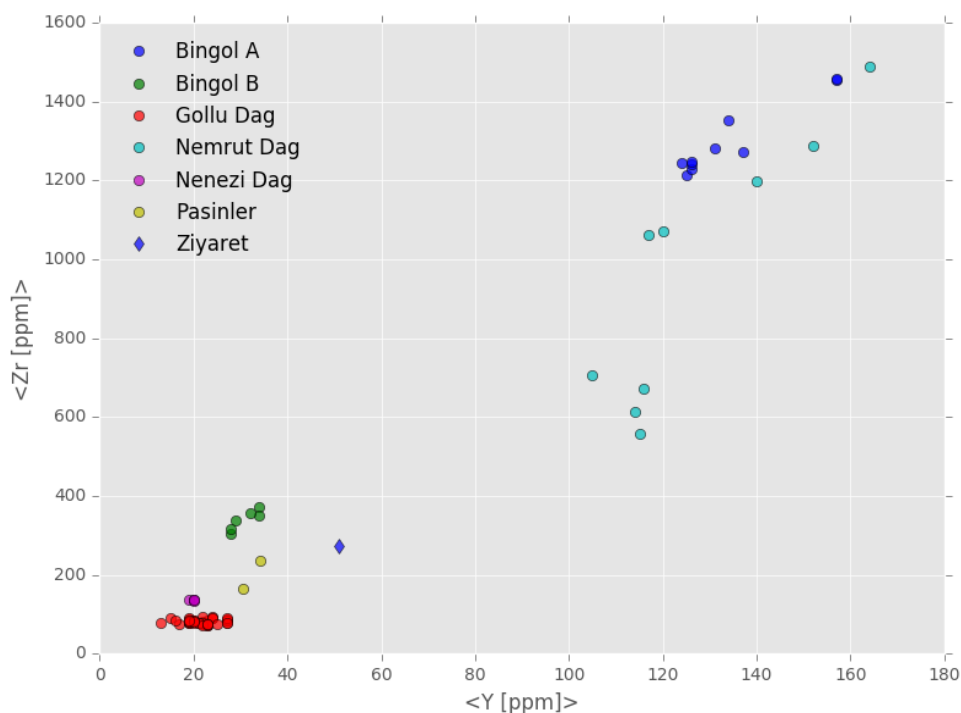


Fig. 4.2. Representación bidimensional Y [ppm] – Zr [ppm]

Por lo tanto para el cálculo de los diferentes subgrupos se utilizó la función Clustering K-means del programa creando un solo conglomerado para las tres nuevas fuentes. En la Figura 4. 3 se observa la representación con los nuevos conglomerados formados.

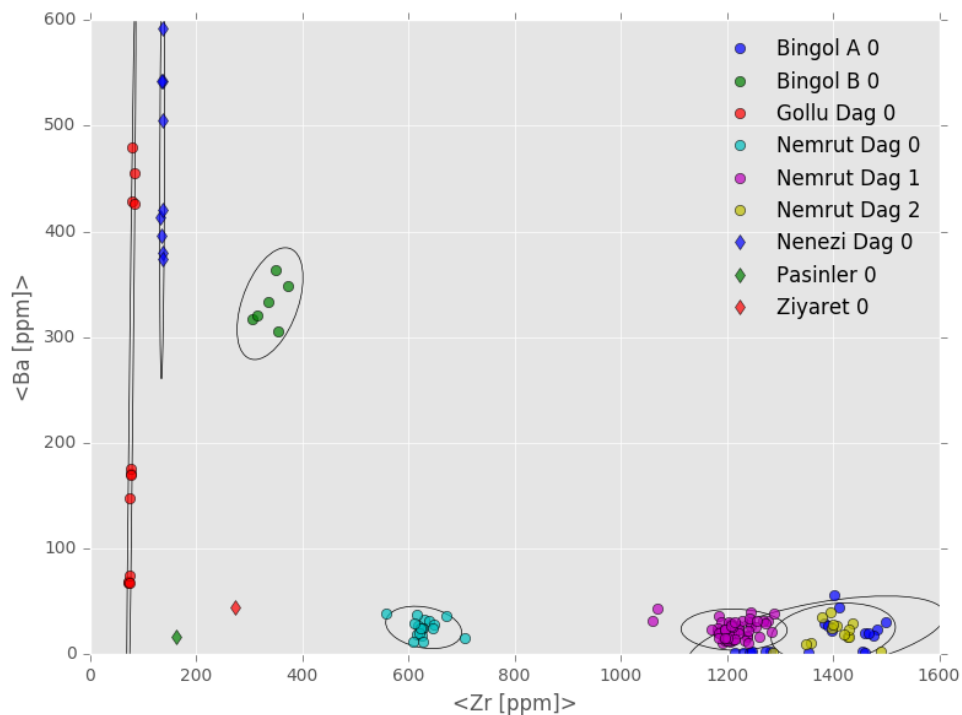


Fig. 4.3. Representación bidimensional Zr [ppm] – Ba [ppm]

4.3. Mejora del programa

Como ya se ha dicho anteriormente, no era necesario implementar ningún algoritmo nuevo ni ningún tratamiento estadístico de los datos adicional. Las nuevas mejoras del programa están destinadas a observar los resultados proporcionados por el análisis de procedencia de las diferentes muestras de obsidiana.

4.3.1. Random Forest

El Random Forest es un algoritmo que consiste en una combinación de árboles de decisión, dependiendo cada uno de ellos de una serie de variables aleatorias. La predicción final viene dada por la moda, el valor que más se repite, de las predicciones de cada uno de los árboles individualmente.

Con el propósito de mejorar el análisis de procedencia de obsidiana se implementó el cálculo del número de veces que una muestra se ha predicho en una fuente. De esta manera, al ejecutar el algoritmo, se obtiene una tabla resumen que muestra el número de veces que dicho algoritmo asigna la muestra a la fuente en cuestión (véase Figura 4.4).

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Pasinler 0	Ziyaret 0	Nulo
Mostra 1	15	0	451	0	14	2	11	7	0	0
Mostra 2	9	31	344	1	55	1	50	6	3	0
Mostra 3	12	0	462	0	18	1	4	3	0	0
Mostra 4a	10	0	358	29	47	2	0	52	2	0
Mostra 4b	14	0	382	19	48	0	0	36	1	0
Mostra 5	16	0	406	17	28	0	0	33	0	0
Mostra 6	10	0	401	20	39	1	0	28	1	0
Mostra 7	22	0	334	32	60	3	0	49	0	0

Fig. 4.4. Tabla resumen para los cálculos del Random Forest

Utilizando la nueva tabla se puede observar cuantos árboles de decisión han asignado la muestra a la fuente predicha y decidir si este valor es significativo o no. En caso de que dicho valor fuera parecido al de otra fuente se procedería a un análisis individual únicamente con esas dos fuentes.

Esta nueva herramienta permite realizar análisis de procedencia mucho más precisos. Anteriormente solo se podía observar la predicción final del algoritmo, en cambio, ahora, el usuario puede observar como se ha producido esta decisión final.

4.3.2. Selección de variables

Una vez realizada la implementación de la tabla resumen del Random Forest se procedió a realizar una mejora en el análisis individual de cada muestra. Para ello se añadieron dos pestañas nuevas para “Seleccionar” y “Deseleccionar” variables en la interfaz del Random Forest, tal como muestra la Figura 4.5.

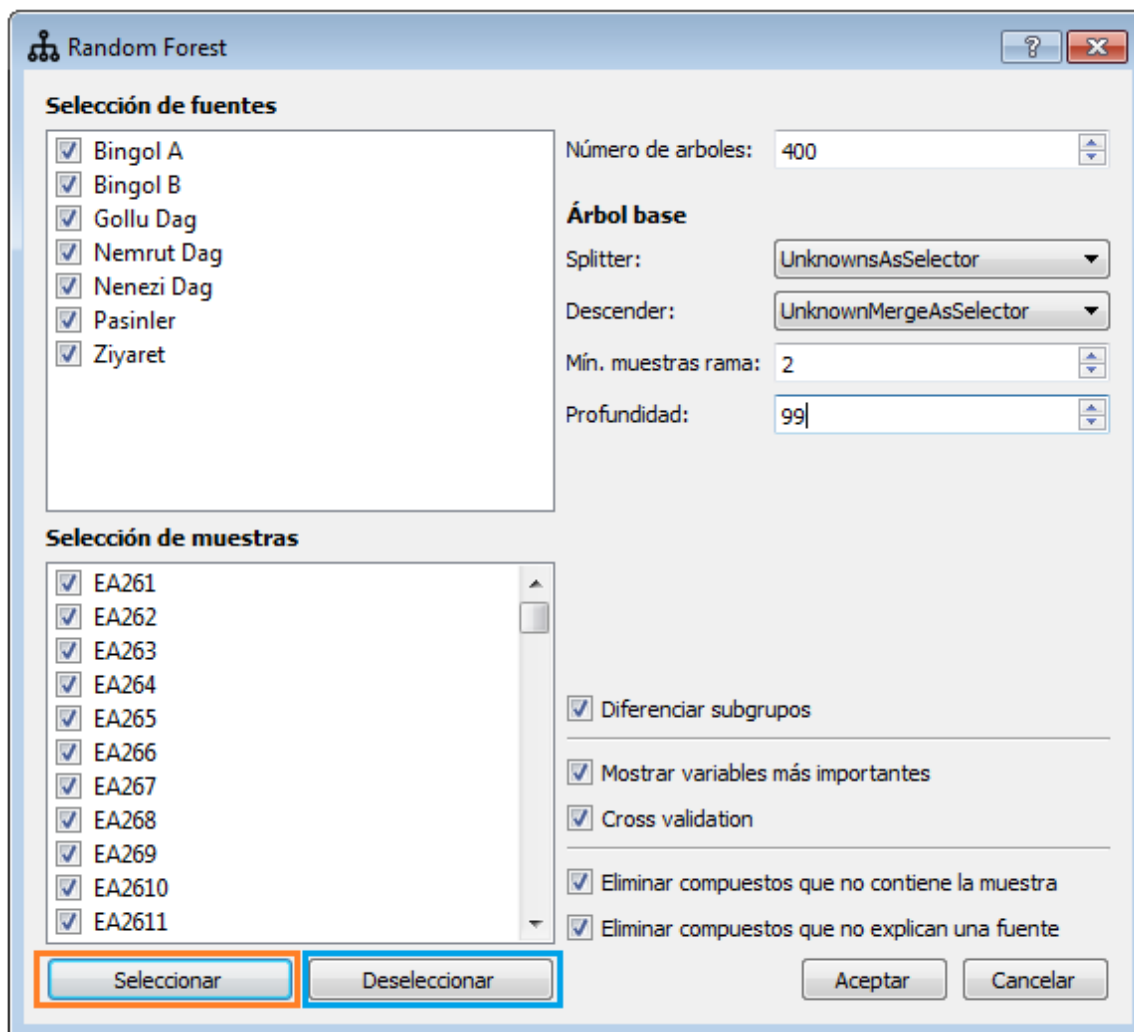


Fig. 4.5. Diálogo Random Forest

Gracias a esta nueva función, en caso de querer analizar una sola muestra de un gran conjunto de muestras se puede realizar un Random Forest de dicha muestra rápidamente, deseleccionando todas las muestras y seleccionando aquella que se requiera observar.

Del mismo modo que en el Random Forest, también se añadió esta nueva función en la interfaz del graficado (véase Figura 4.6).

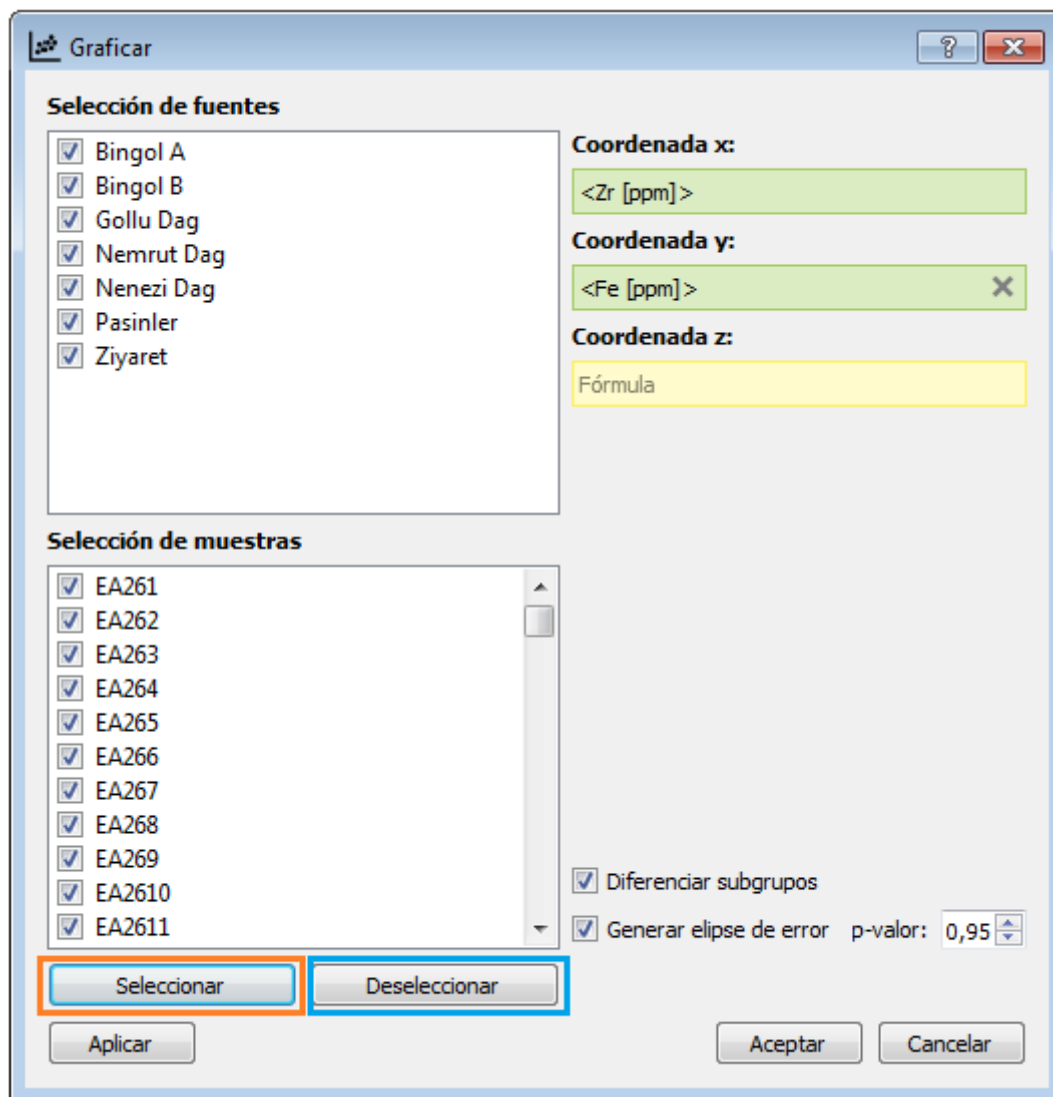


Fig. 4.6. Diálogo de representación gráfica

4.4. Ejecutable

El objetivo principal de crear un ejecutable es utilizar el programa de manera rápida y sencilla sin necesidad de acceder a Python ni escribir ningún comando previo con el fin de que cualquier persona pueda utilizar este programa.

Para realizar el ejecutable del programa se estudiaron diferentes metodologías y programas existentes en Python:

- Py2exe
- PyInstaller
- cx_Freeze

Py2exe

Py2exe es una extensión de Python que convierte scripts de Python en programas ejecutables de Windows. Dichos programas son capaces de funcionar sin requerir una instalación previa de Python.

PyInstaller

PyInstaller es un programa que convierte los programas de Python en ejecutables autónomos para Windows, Linux, Mac OS X, etc. Las principales ventajas de PyInstaller son que trabaja tanto con Python 2.7 como con Python 3.3-3.5, crea ejecutables pequeños y utiliza el soporte del sistema operativo para cargar las bibliotecas dinámicas, garantizando así la compatibilidad con el sistema operativo.

cx_Freeze

cx_Freeze es un conjunto de scripts y módulos para convertir scripts de Python en ejecutables, del mismo modo que lo realiza Py2exe. A diferencia de este, cx_Freeze es multiplataforma, es decir, funciona con cualquier plataforma en la que Python trabaje. Es compatible tanto con Python 2.7 como con Python 3.3.

Sin embargo, a la hora de llevar a cabo la creación de un ejecutable con cualquiera de estos programas fue imposible su realización. Algunos de los principales motivos de los errores fueron la existencia de bugs en algunos de los instaladores y la utilización de un gran número de librerías como por ejemplo la librería Orange.

Orange es una librería de Python orientada a la minería de datos. Posee una gran ventaja y es que permite la posibilidad de trabajar con valores perdidos. Sin embargo, Orange es una librería de Python que lleva tiempo sin actualizarse produciendo problemas con la creación de dichos ejecutables.

Para solucionar la imposibilidad de crear un ejecutable para Windows se abordó el problema desde otra perspectiva. Con tal de cumplir el objetivo y que cualquier persona pudiera utilizar el programa sin necesidad de conocer la programación en Python, se decidió crear un pequeño programa que funcionara como ejecutable, introduciendo todas las órdenes necesarias en Python para abrir el Source Seeker. Para la creación de este nuevo ejecutable se utilizaron los llamados archivos batch.

Archivo batch

Un archivo batch es un archivo de procesamiento por lotes. Se trata de archivos de texto guardados con la extensión .bat que contienen un conjunto de instrucciones. Las instrucciones que contienen estos archivos son órdenes a ejecutar en un intérprete de órdenes. Cuando se ejecuta este archivo, un programa shell, el command de Windows en nuestro caso, lee el archivo y ejecuta las órdenes en grupo, de forma secuencial, permitiendo automatizar diversas tareas.

De este modo, se creó un archivo batch[16] llamado “*Ejecutable.bat*” que contenía todas las instrucciones necesarias para abrir el programa Source Seeker sin necesidad de entrar en Python. Cuando el archivo batch es ejecutado, este abre el command de Windows y inicializa el programa Source Seeker (véase Figura 4.7).

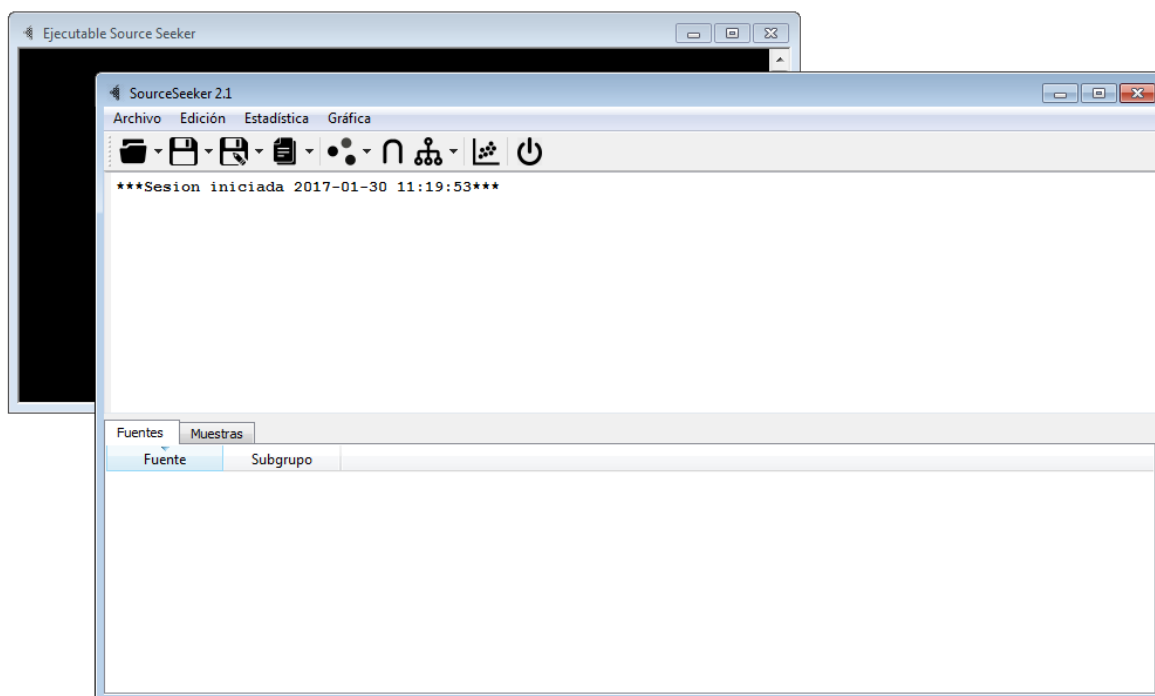


Fig. 4.7. Ejecutable e interfaz principal del programa

Finalmente, para concluir con el ejecutable y mediante el programa llamado bat_to_exe se transformó el programa en batch[15] a un programa .exe llamado SourceSeeker.exe con el icono del programa tal y como indica la Figura 4.8.



Fig. 4.8. Nuevo ejecutable del programa

4.5. Instalador

Para poder ejecutar el programa se requiere tener previamente instalado Python. No obstante, el programa Source Seeker necesita disponer de la instalación de una serie de módulos de Python para poder funcionar. Estos disponen de programas para realizar los cálculos, graficar, hacer el tratamiento de los datos, etc. Los módulos o librerías necesarias son los siguientes:

- Scikit-learn
- Numpy
- Scipy
- PyQt4
- Orange
- Matplotlib
- xlrd
- xlwt

Para hacer más fácil la instalación de esta serie de módulos se creó un pequeño programa en batch, de la misma forma que el ejecutable Source Seeker, llamado Instalador.bat. Al ejecutar el instalador se abre una ventana (véase Figura 4.9) mediante la cual, dependiendo de la versión de Windows que se seleccione, instalará todos los módulos necesarios, descritos anteriormente, de forma automática.

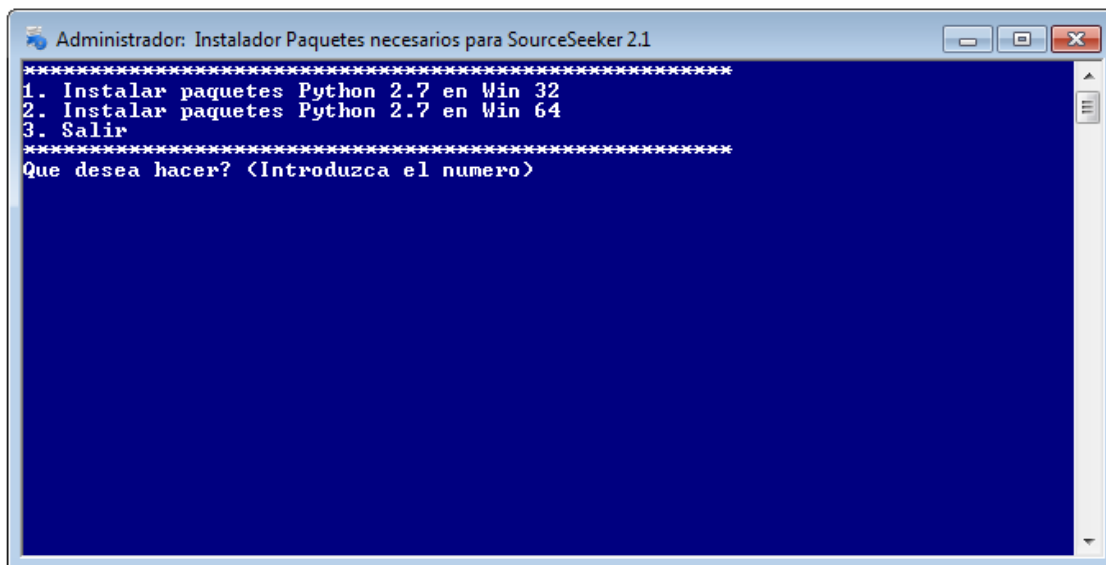


Fig. 4.9. Interfaz del Instalador de módulos de Python

La instalación de dichos módulos se realiza mediante un programa de Python llamado get-pip. El instalador instala este programa de forma automática y secuencialmente instala los módulos sin que el usuario necesite conocimientos de lenguaje en Python.

Para concluir con el instalador, se transformó el archivo .bat a un archivo .exe tal y como se ha explicado anteriormente.

Finalmente, para facilitar la posible distribución del programa, ahorrando espacio de almacenamiento y tiempo de transmisión al enviar estos archivos, se creó una carpeta comprimida mediante el programa WinRAR[23].

La carpeta comprimida contiene los siguientes archivos, tal y como muestra la Figura 4.10:

- Instalador de paquetes.
- Programa Source Seeker.
- Instalador de Python.
- Ejecutable Source Seeker.

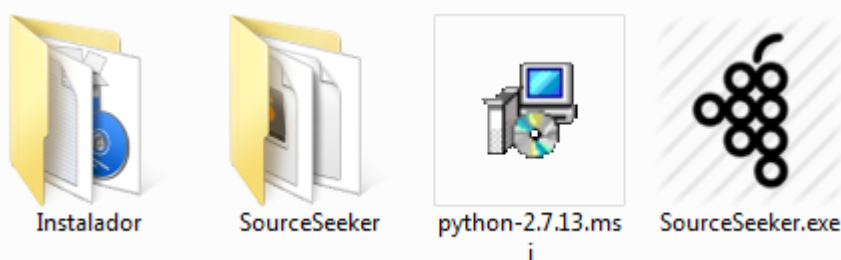


Fig. 4.10. Archivos contenidos en la carpeta comprimida

5. Análisis de los artículos

Con el objetivo de validar el funcionamiento del programa y observar las posibles deficiencias que presenta se preparó una batería de artículos los cuales contienen muestras de obsidiana analizadas químicamente y asignadas cada una a sus posibles fuentes de origen.

Aunque en los dos anteriores proyectos ya se realizaron estudios de procedencia con el programa, el estudio planteado a continuación es mucho más exhaustivo, ya que contiene un total de 10 artículos, con 402 muestras en total para analizar.

Dicho estudio de procedencia se ejecutará con el programa Source Seeker y las mejoras realizadas en el transcurso del presente proyecto, con lo cual dispondremos de un mayor número de fuentes de origen y una tabla de resultados de la función Random Forest.

5.1. Cálculos previos

En la base de datos del programa algunos de los elementos químicos se encuentran expresados en forma de óxido mientras que otros están expresados como elementos puros. Por este motivo, antes de realizar el análisis de procedencia hay que expresar cada uno de los elementos químicos que forman las muestras de obsidiana en las mismas unidades que están expresados los elementos químicos de las fuentes.

Los elementos expresados en forma de óxido en los artículos y que necesitan expresarse en forma de compuesto puro son el hierro (Fe), el manganeso (Mn), el calcio (Ca) y el titanio (Ti). A continuación se muestran los factores de conversión que permiten expresar los óxidos en elementos puros.

- FeO [wt%] a Fe [ppm]:

$$x [\% FeO] \cdot \frac{1 \text{ mol } FeO}{71,844 \text{ g } FeO} \cdot \frac{1 \text{ mol } Fe}{1 \text{ mol } FeO} \cdot \frac{55,845 \text{ g } Fe}{1 \text{ mol } Fe} \cdot \frac{10000 \text{ [ppm]}}{1 [\% FeO]} \quad (\text{Ec.4.1})$$

- Fe₂O₃ [wt%] a Fe [ppm]:

$$x [\% Fe_2O_3] \cdot \frac{1 \text{ mol } Fe_2O_3}{159,69 \text{ g } Fe_2O_3} \cdot \frac{2 \text{ mol } Fe}{1 \text{ mol } Fe_2O_3} \cdot \frac{55,845 \text{ g } Fe}{1 \text{ mol } Fe} \cdot \frac{10000 \text{ [ppm]}}{1 [\% Fe_2O_3]} \quad (\text{Ec.4.2})$$

- MnO [wt%] a Mn [ppm]:

$$x [\% MnO] \cdot \frac{1 \text{ mol } MnO}{70,937 \text{ g } MnO} \cdot \frac{1 \text{ mol } Mn}{1 \text{ mol } MnO} \cdot \frac{54,938 \text{ g } Mn}{1 \text{ mol } Mn} \cdot \frac{10000 \text{ [ppm]}}{1 [\% MnO]} \quad (\text{Ec.4.3})$$

- Mn_3O_4 [wt%] a Mn [ppm]:

$$x [\% MnO] \cdot \frac{1 \text{ mol } Mn_3O_4}{228,812 \text{ g } Mn_3O_4} \cdot \frac{3 \text{ mol } Mn}{1 \text{ mol } Mn_3O_4} \cdot \frac{54,938 \text{ g } Mn}{1 \text{ mol } Mn} \cdot \frac{10000 [ppm]}{1 [\% Mn_3O_4]} \quad (\text{Ec.4.4})$$

- CaO [wt%] a Ca [ppm]:

$$x [\% CaO] \cdot \frac{1 \text{ mol } CaO}{56 \text{ g } CaO} \cdot \frac{1 \text{ mol } Ca}{1 \text{ mol } CaO} \cdot \frac{40 \text{ g } Ca}{1 \text{ mol } Ca} \cdot \frac{10000 [ppm]}{1 [\% CaO]} \quad (\text{Ec.4.5})$$

- TiO_2 [wt%] a Ti [ppm]:

$$x [\% TiO_2] \cdot \frac{1 \text{ mol } TiO_2}{79,867 \text{ g } TiO_2} \cdot \frac{1 \text{ mol } Ti}{1 \text{ mol } TiO_2} \cdot \frac{47,867 \text{ g } Ti}{1 \text{ mol } Ti} \cdot \frac{10000 [ppm]}{1 [\% TiO_2]} \quad (\text{Ec.4.6})$$

5.2. Localización de los yacimientos

A continuación, tal y como se ha explicado anteriormente, se presenta el análisis de procedencia de las muestras de obsidiana contenidas en los artículos proporcionados por el director del proyecto. A causa de la gran homogeneidad de las muestras, casi todas contienen los mismos elementos, el análisis de procedencia se puede aplicar sobre todo el conjunto de muestras de cada artículo al mismo tiempo.

Para realizar el análisis se descarta utilizar el árbol de decisiones, ya que era preferible utilizar el algoritmo Random Forest. Este último consiste en un conjunto de árboles de decisiones, obteniendo resultados más precisos y fiables.

En todos y cada uno de los análisis se emplean 400 árboles de decisiones para evitar tiempos de ejecución elevados y obtener buenos resultados. Los parámetros *splitter* y *descender* son configurados como *UnknownsAsSelector* y *UnknownMergeAsSelector* respectivamente, por defecto. Se limita a una profundidad de 99 y se diferencian las muestras entre subgrupos, eliminando los elementos que no contengan las muestras o no expliquen alguna fuente. Por último se habilita el cálculo de las variables más importantes, obteniendo así un grafico que se utilizará para validar el análisis.

5.2.1. Artículo 1

En este artículo[2] se realiza un estudio interno del volcán Göllü Dağ, a causa de la gran variabilidad de muestras que presenta dicho volcán, más de diez tipos diferentes de obsidiana fueron identificados, los métodos empleados en el artículo fueron principalmente la representación bidimensional de concentración de elementos y el análisis de

componentes principales. Se recolectaron muestras de distintas ubicaciones y se concluyó que existían diferentes subgrupos en las muestras del volcán. Una de las posibles causas de dicha variabilidad es el momento de la creación de la obsidiana durante el enfriamiento de la cámara magmática, la forma y los factores involucrados durante el proceso.

Antes de realizar el análisis de las muestras con el programa es necesario convertir las concentraciones de algunos elementos químicos expresados en forma de óxidos en forma de elemento. Mediante las Ec. 4.2 y Ec. 4.5 expresaremos el óxido de hierro y el óxido de calcio en hierro y calcio.

Una vez realizadas las conversiones pertinentes se procede a ejecutar el análisis de procedencia. Para ello, se seleccionan todas las muestras y fuentes disponibles y se ejecuta el algoritmo Random Forest. En la Figura 5.1 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
Mostra 1	Gollu Dag 0
Mostra 2	Gollu Dag 0
Mostra 3	Gollu Dag 0
Mostra 4a	Gollu Dag 0
Mostra 4b	Gollu Dag 0
Mostra 5	Gollu Dag 0
Mostra 6	Gollu Dag 0
Mostra 7	Gollu Dag 0

AP	CA	Sensitivity	Specificity
0.522	0.763	1.000	1.000

Fig. 5.1. Resultados de predicción y validación Random Forest

En la Figura 5.2 se observa el resultado del cálculo de las variables más importantes.

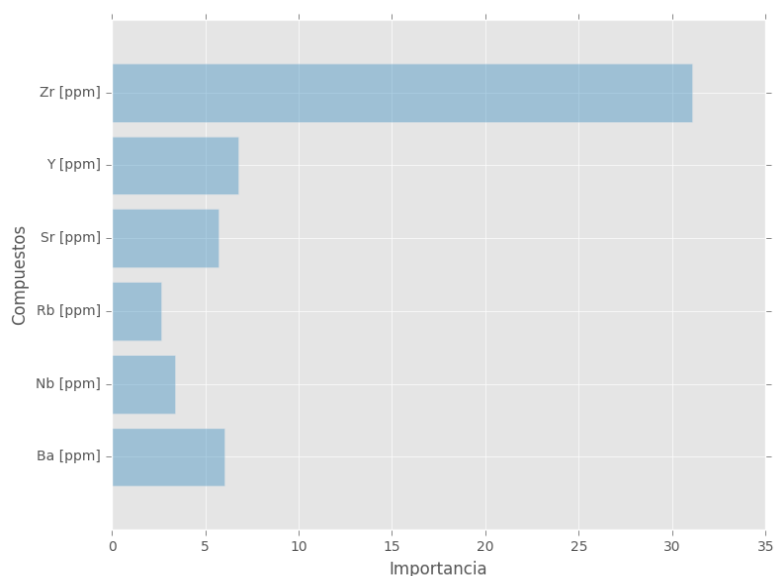


Fig. 5.2. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.2, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos, Y [ppm] – Zr [ppm] (Figura 5.3) y Y [ppm] – Zr [ppm] – Ba [ppm] (Figura 5.4).

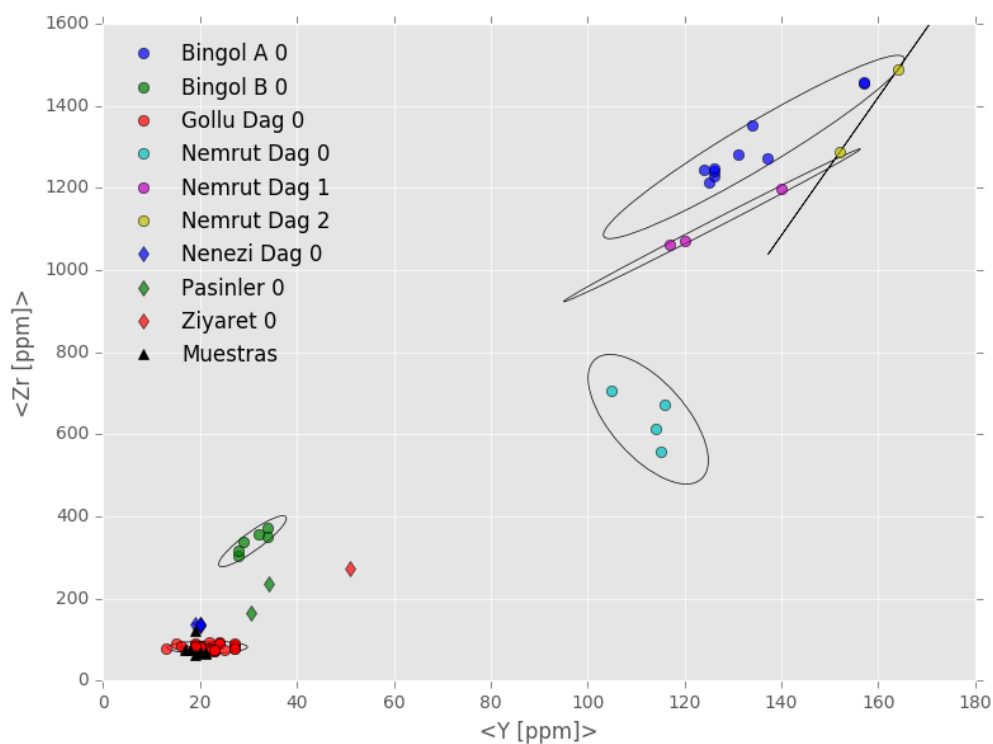


Fig. 5.3. Representación bidimensional Y [ppm] – Zr [ppm]

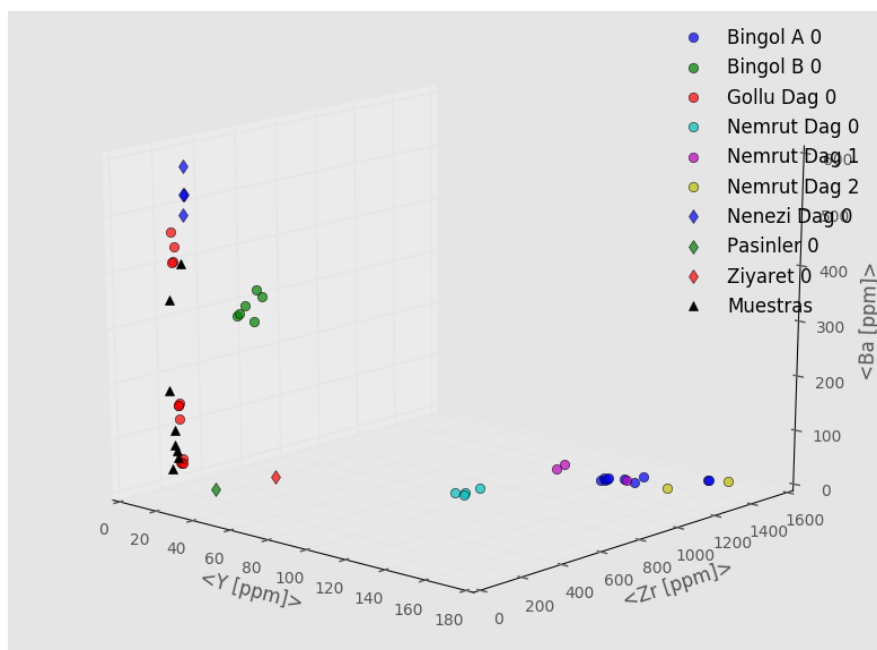


Fig. 5.4. Representación tridimensional Y [ppm] – Zr [ppm] – Ba [ppm]

Observando bien las representaciones y la predicción del Random Forest podemos concluir que las muestras analizadas pertenecen a la fuente Göllü Dağ 0. En los gráficos se observa cómo se distingue bien la distribución de las muestras, sin ningún tipo de ambigüedad. Los resultados obtenidos coinciden con las conclusiones extraídas en este artículo, asignando todas estas muestras a la fuente Göllü Dağ.

Por otro lado, si nos fijamos en la representación tridimensional, Figura 5.4, se aprecian tres subgrupos en la fuente Göllü Dağ y dichas muestras analizadas se distribuyen en estos subgrupos. El presente artículo señala que la fuente Göllü Dağ presenta diferentes subgrupos de muestras, no obstante, debido a la poca cantidad de datos que dispone el programa, no es posible realizar subgrupos de la presente fuente con la función Clustering K-means.

5.2.2. Artículo 2

El objetivo principal de este estudio[3] es demostrar que la obsidiana proveniente de la fuente Dasht-i-Nawur, situada en Afganistán, es fácilmente distinguible de otras fuentes de obsidiana de Oriente Próximo y el suroeste de Asia, debido a la existencia de sistemas volcánicos con obsidiana desde la zona del Mar Caspio hasta el Asia Central, de los que no se extrajo la obsidiana en Oriente Próximo pero sí, probablemente, la utilizada en Oriente Medio, como sería el caso de yacimientos arqueológicos afganos. Para ello, en el artículo estudiado se analizaron los componentes principales de diferentes muestras de obsidiana y se realizaron representaciones bidimensionales. Los resultados extraídos del análisis indicaban que la obsidiana proveniente de esta fuente exhibe una composición química

única que la diferencia de todas las otras fuentes de obsidiana caracterizadas.

En el presente artículo no es necesario realizar ninguna conversión, por lo tanto se analizan directamente las muestras proporcionadas. Para ello se seleccionan todas las muestras y fuentes y se ejecuta el algoritmo. En la Figura 5.5 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
DASH-1	Nenezi Dag 0
DASH-2	Nenezi Dag 0
DASH-3	Nenezi Dag 0
DASH-5	Nenezi Dag 0
DASH-6	Nenezi Dag 0
DASH-7	Nenezi Dag 0

AP	CA	Sensitivity	Specificity
0.674	0.903	1.000	1.000

Fig 5.5. Resultados predicción y validación Random Forest

En la Figura 5.6 se observa el resultado del cálculo de las variables más importantes.

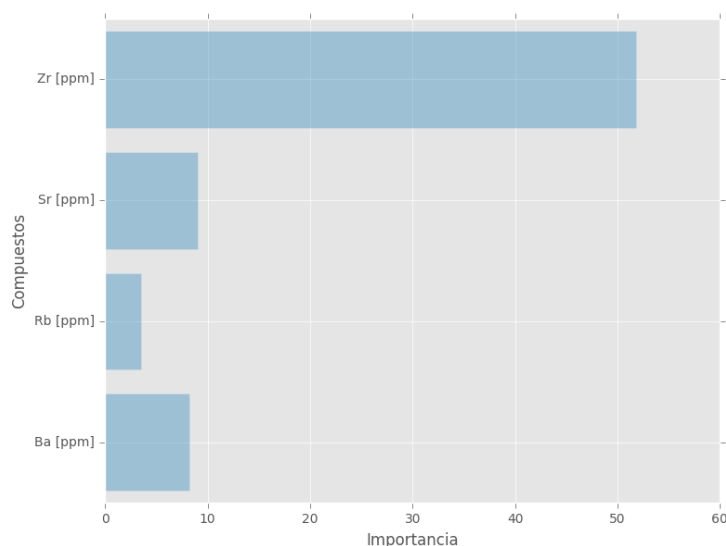


Fig. 5.6. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.6, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Sr [ppm] – Zr [ppm] (Figura 5.7) y Sr [ppm] – Zr [ppm] – Rb [ppm] (Figura 5.8).

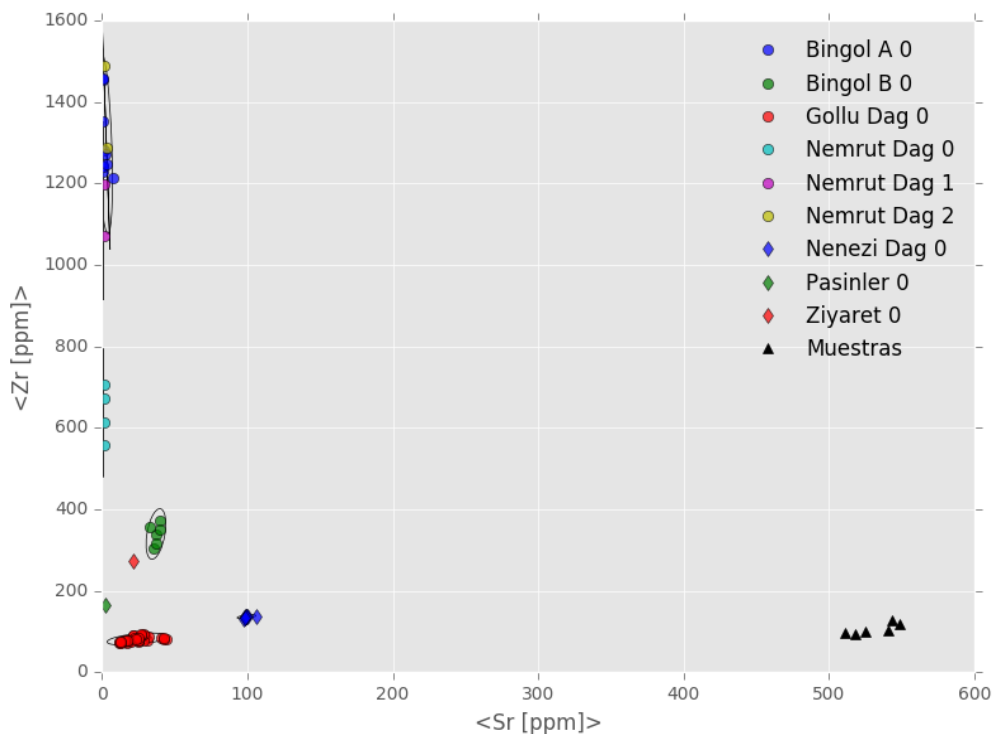


Fig. 5.7. Representación bidimensional Sr [ppm] – Zr [ppm]

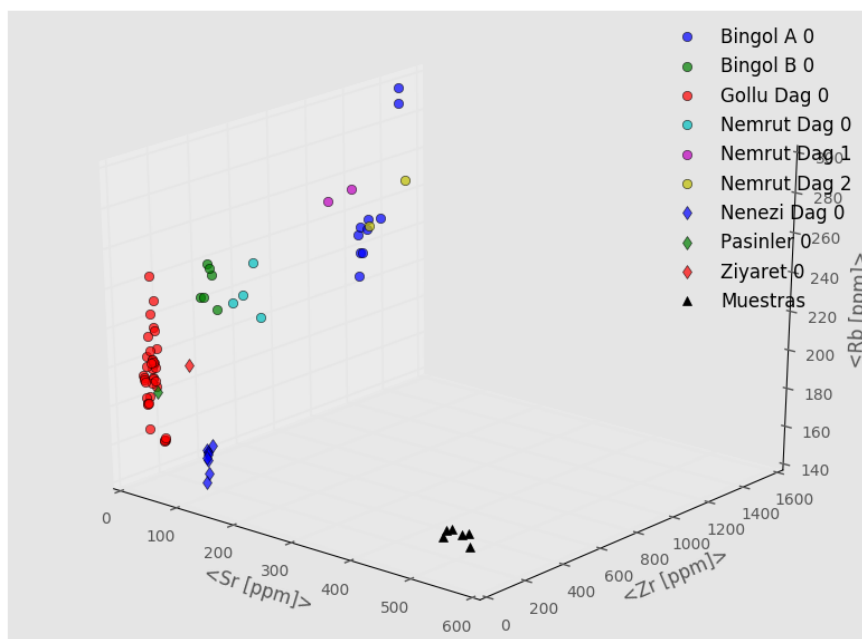


Fig. 5.8. Representación tridimensional Sr [ppm] – Zr [ppm] – Rb [ppm]

Las muestras de este artículo provienen de la fuente Dasht-i-Nawur, una fuente no disponible para nuestro programa. Por este motivo, a la hora de realizar el análisis, el programa asigna las muestras a una de sus fuentes, la que más similitud tiene con las muestras. En este caso son asignadas a la fuente Nenezi Dağ. Si observamos los gráficos representados anteriormente se puede concluir que las muestras no pertenecen a ninguna de las fuentes indicadas.

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Pasinler 0	Ziyaret 0	Nulo
DASH-1	3	0	0	1	22	0	296	78	0	0
DASH-2	2	0	111	0	6	0	179	102	0	0
DASH-3	0	0	111	0	1	0	268	20	0	0
DASH-5	0	0	111	0	2	0	283	4	0	0
DASH-6	0	0	111	0	1	0	268	20	0	0
DASH-7	1	0	0	0	20	0	356	23	0	0

Fig. 5.9. Tabla resumen Random Forest

Si tenemos presente la tabla de árboles, (véase Figura 5.9) podemos ver que no existe una asignación clara, ya que en algunas muestras la mitad de árboles asignan la muestra a una fuente y la otra mitad a otra fuente. Tal y como se ha explicado anteriormente, el programa no puede asignar correctamente la muestra al no disponer de su fuente original.

Del mismo modo que el artículo, se concluye que las muestras de obsidiana pertenecientes a esta fuente presentan una composición química única, hecho que permite distinguir las del resto de fuentes.

5.2.3. Artículo 3

En este tercer artículo[4] se estudia obsidiana proveniente de Tell Kurdu, al sur de Turquía, utilizando los métodos ya descritos, la representación bidimensional. A parte de la tradicional asignación a la Capadocia (Göllü Dağ) y el este de Anatolia (Bingöl i Nemrut Dağ), los resultados del estudio señalan nuevas fuentes situadas en la zona del lago Van (Ziyaret) y noreste de Anatolia (Pasinler).

Antes de realizar el análisis pertinente, efectuaremos la conversión de los compuestos en forma de óxido Fe_2O_3 , MnO , CaO y TiO_2 a su forma pura Fe , Mn , Ca y Ti mediante las ecuaciones Ec. 4.2, Ec. 4.3, Ec. 4.5, y Ec. 4.6.

Una vez realizadas las conversiones se ejecuta el algoritmo Random Forest. En la Figura 5.10 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
TK3489	Gollu Dag 0
TK4032	Gollu Dag 0
TK4918	Gollu Dag 0
TK3044	Ziyaret 0
TK4082	Pasinler 0
TK4377	Bingol B 0
TK3265	Bingol A 0
TK4444	Bingol A 0
TK4969	Bingol A 0

AP	CA	Sensitivity	Specificity
0.522	0.763	1.000	1.000

Fig. 5.10. Resultados predicción y validación Random Forest

En la Figura 5.11 se observa el resultado del cálculo de las variables más importantes.

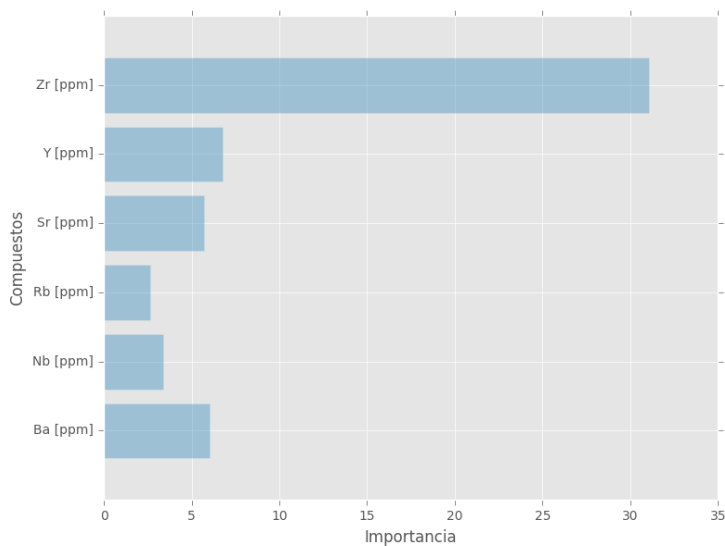


Fig. 5.11. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.11, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Y [ppm] – Rb [ppm] (Figura 5.12) y Y [ppm] – Zr [ppm] – Rb [ppm] (Figura 5.13).

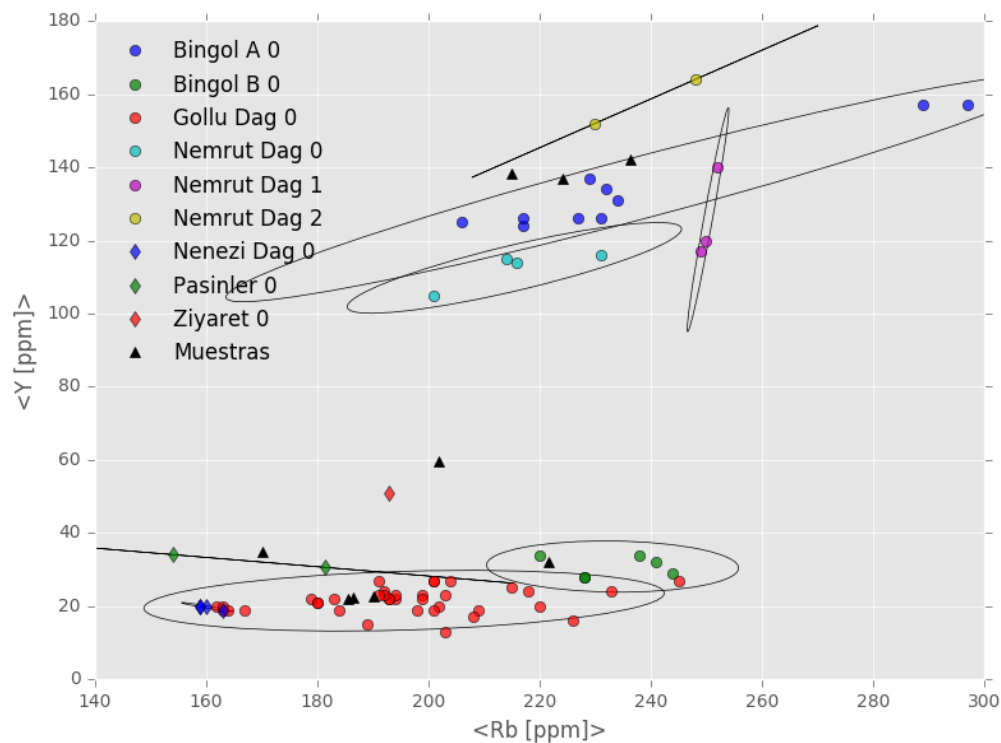


Fig. 5.12. Representación bidimensional Y [ppm] – Rb [ppm]

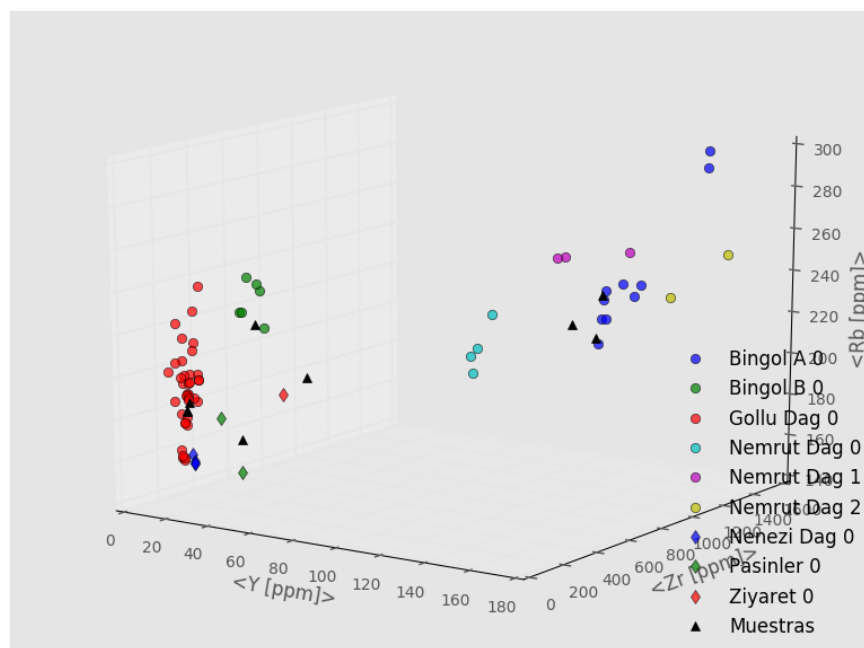


Fig. 5.13. Representación tridimensional Y [ppm] – Zr [ppm] – Rb [ppm]

Al observar los gráficos podemos concluir que las muestras referentes a las fuentes Göllü Dağ, Bingöl B, Ziyaret y Pasinler son asignadas correctamente. Las tres muestras restantes son asignadas a la fuente Bingöl A, tal y como se observa en las Figuras 5.12 y 5.13.

El artículo estudiado sitúa estas últimas muestras en dos posibles regiones: la primera es el área de Bingöl y la segunda es el volcán Nemrut Dağ. En la representación tridimensional de la Figura 5.13 dichas muestras están en la zona de Bingöl A, bastante alejadas de la fuente Nemrut Dağ, por lo tanto podemos concluir que las muestras probablemente proceden de Bingöl A.

5.2.4. Artículo 4

Este estudio[6] es similar a los anteriores aunque contiene más muestras por identificar localizadas en la parte central de Anatolia, en el yacimiento de Çatalhöyük. En el artículo publicado, la composición química de las muestras las relaciona con dos fuentes geológicas de la zona sur de la Capadocia: Göllü Dağ y Nenezi Dağ. La distinción entre las dos fuentes se realiza mediante una representación bidimensional, Zn [ppm] – Zr [ppm].

En el presente artículo no es necesario realizar ninguna conversión, por lo tanto se analizan directamente las muestras proporcionadas. Para ello se seleccionan todas las muestras y fuentes y se ejecuta el algoritmo. En la Figura 5.14 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion		
OB251	Nenezi Dag 0	OB273	Gollu Dag 0
OB252	Nenezi Dag 0	OB274	Gollu Dag 0
OB253	Gollu Dag 0	OB275	Gollu Dag 0
OB254	Gollu Dag 0	OB276	Gollu Dag 0
OB255	Gollu Dag 0	OB277	Nenezi Dag 0
OB256	Nenezi Dag 0	OB278	Nenezi Dag 0
OB257	Gollu Dag 0	OB279	Nenezi Dag 0
OB258	Gollu Dag 0	OB280	Nenezi Dag 0
OB259	Gollu Dag 0	OB281	Gollu Dag 0
OB260	Gollu Dag 0	OB282	Nenezi Dag 0
OB261	Nenezi Dag 0	OB283	Nenezi Dag 0
OB262	Nenezi Dag 0	OB284	Gollu Dag 0
OB263	Gollu Dag 0	OB285	Gollu Dag 0
OB264	Gollu Dag 0	OB286	Nenezi Dag 0
OB265	Nenezi Dag 0	OB287	Nenezi Dag 0
OB266	Gollu Dag 0	OB288	Gollu Dag 0
OB267	Nenezi Dag 0	OB289	Nenezi Dag 0
OB268	Gollu Dag 0	OB290	Gollu Dag 0
OB269	Gollu Dag 0	OB291	Gollu Dag 0
OB270	Nenezi Dag 0	OB292	Gollu Dag 0
OB271	Gollu Dag 0	RGM1-H1 1	Nenezi Dag 0
OB272	Gollu Dag 0	RGM1-H1 2	Nenezi Dag 0
		RGM1-H1 3	Nenezi Dag 0
		RGM1-H1 4	Nenezi Dag 0

AP	CA	Sensitivity	Specificity
0.437	0.683	1.000	1.000

Fig. 5.14. Resultados predicción y validación Random Forest

En la Figura 5.15 se observa el resultado del cálculo de las variables más importantes.

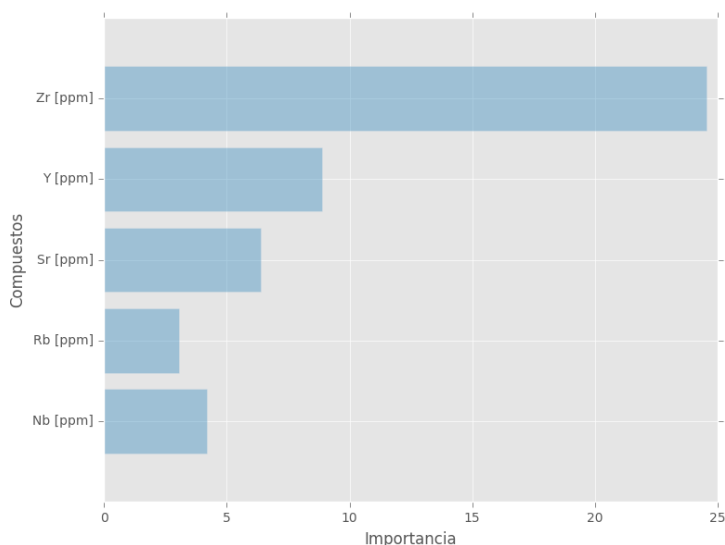


Fig. 5.15. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.15, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Y [ppm] – Zr [ppm] (Figura 5.16) y Y [ppm] – Zr [ppm] – Sr [ppm] (Figura 5.17).

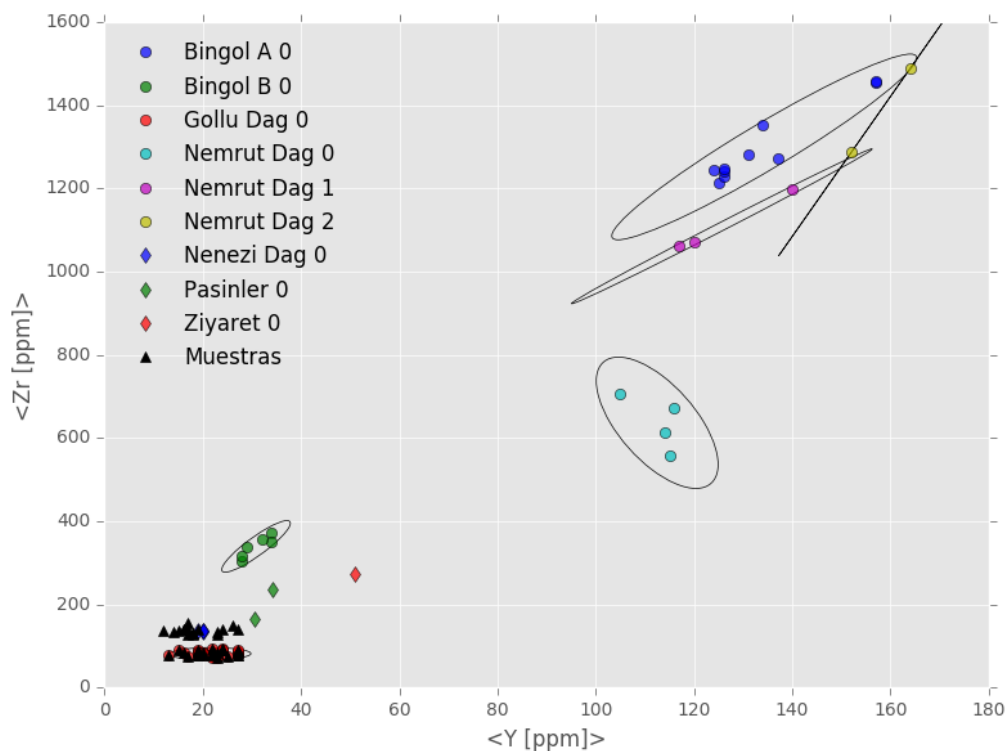


Fig. 5.16. Representación bidimensional Y [ppm] – Zr [ppm]

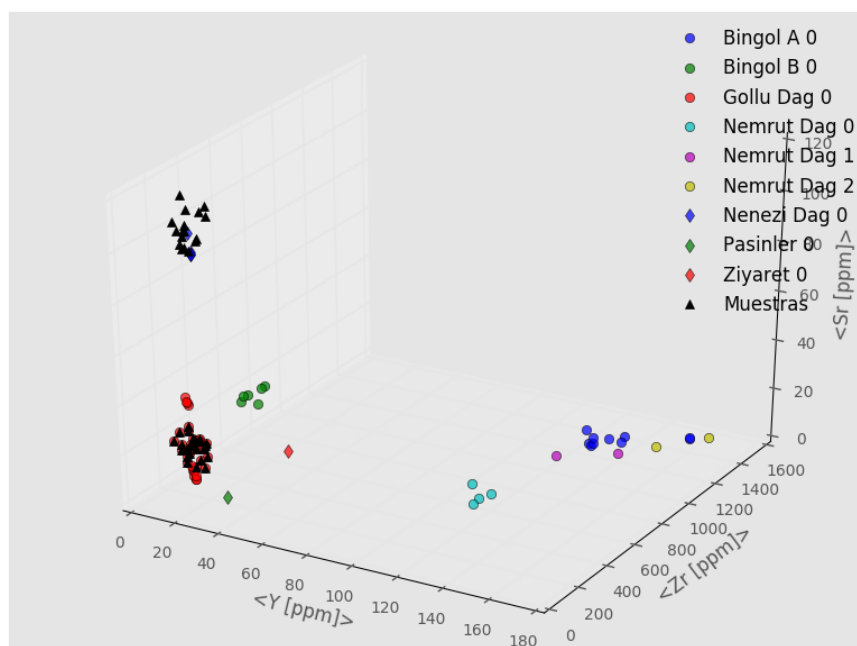


Fig. 5.17. Representación tridimensional Y [ppm] – Zr [ppm] – Sr [ppm]

En este presente análisis, las muestras son asignadas a las fuentes Göllü Dağ y Nenezi Dağ. En los anteriores gráficos se puede observar cómo se distinguen perfectamente las muestras pertenecientes a cada volcán. Si nos fijamos en la Figura 5.14, existen cuatro muestras añadidas en el artículo y catalogadas como “Standard”, estas muestras no pertenecen a ninguna fuente en concreto y por ese motivo no han sido representadas en las Figuras 5.16 y 5.17.

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Muestra	Gollu Dag 0	Nenezi Dag 0	Nulo
OB251	0	4	131	0	66	0	196	OB251	88	312	0
OB252	0	1	133	0	51	0	215	OB252	85	315	0
OB256	0	11	138	0	54	0	197	OB256	85	315	0
OB261	0	3	127	0	73	0	194	OB261	86	314	0
OB262	0	3	112	0	69	0	214	OB262	78	322	0
OB265	0	4	127	0	52	0	217	OB265	90	310	0
OB267	0	0	75	0	39	0	286	OB267	56	344	0
OB270	0	7	135	0	55	0	203	OB270	87	313	0
OB278	0	1	133	0	52	0	214	OB278	83	317	0
OB279	0	0	98	0	38	0	264	OB279	64	336	0
OB280	0	4	127	0	46	0	223	OB280	85	315	0
OB282	0	0	1	0	78	0	314	OB282	16	384	0
OB283	0	2	124	0	54	0	220	OB283	87	313	0
OB286	0	4	130	0	53	0	213	OB286	87	313	0
OB287	0	3	113	0	41	0	243	OB287	75	325	0

Fig. 5.18. Tablas resumen Random Forest

Si nos fijamos en la tabla resumen que proporciona el programa las muestras referentes a la fuente Nenezi Dağ tienen un número de asignaciones parecido al de la fuente Göllü Dağ. Sin embargo, haciendo un análisis individual de dichas muestras y seleccionando únicamente las fuentes Göllü Dağ y Nenezi Dağ, se puede observar como dichas muestras pertenecen a la segunda fuente. En la Figura 5.18 se pueden observar las dos tablas resumen del Random Forest.

El artículo en cuestión señala que las muestras provienen de la fuente Göllü Dağ y Nenezi Dağ. Las asignaciones de la investigación coinciden con las conclusiones extraídas por el programa.

5.2.5. Artículo 5

El presente artículo[5] estudia muestras procedentes del este de Anatolia, del yacimiento de Körtik Tepe, provenientes de las fuentes Bingöl A/B y Nemrut Dağ. Para el estudio se realizan representaciones bidimensionales de elementos y de razones de elementos. En el momento en que se encuentra una representación que agrupe bien las muestras de las fuentes de procedencia, se superponen las de los artefactos estudiados y, según si caen en una región u otra, se le asigna a uno u otro volcán.

Ya que no es necesario realizar ningún tipo de cambio de unidad seleccionamos todas las muestras y fuentes disponibles y ejecutamos el programa. En el listado de muestras del presente artículo existían concentraciones negativas, estas concentraciones han sido substituidas por 0, en lugar del valor negativo. En la Figura 5.19 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion	KT.038	Bingol B 0	KT.085	Bingol B 0	KT.039	Bingol A 0	KT.040	Nemrut Dag 1
KT.002	Bingol B 0	KT.041	Bingol B 0	KT.089	Bingol B 0	KT.043	Nemrut Dag 1	KT.044	Bingol A 0
KT.004	Bingol B 0	KT.042	Bingol B 0	KT.090	Bingol B 0	KT.058	Nemrut Dag 1	KT.045	Bingol A 0
KT.005	Bingol B 0	KT.046	Nenezi Dag 0	KT.091	Bingol B 0	KT.059	Bingol A 0	KT.052	Nemrut Dag 1
KT.006	Bingol B 0	KT.047	Bingol B 0	KT.095	Bingol B 0	KT.061	Nemrut Dag 1	KT.053	Nemrut Dag 1
KT.007	Bingol B 0	KT.048	Bingol B 0	KT.097	Bingol B 0	KT.062	Nemrut Dag 1	KT.069	Bingol A 0
KT.008	Bingol B 0	KT.049	Bingol B 0	KT.099	Bingol B 0	KT.065	Nemrut Dag 1	KT.071	Nemrut Dag 1
KT.009	Bingol B 0	KT.050	Bingol B 0	KT.101	Bingol B 0	KT.073	Nemrut Dag 1	KT.092	Nemrut Dag 1
KT.013	Bingol B 0	KT.051	Bingol B 0	KT.102	Bingol B 0	KT.074	Bingol A 0	KT.094	Nemrut Dag 2
KT.014	Bingol B 0	KT.054	Bingol B 0	KT.104	Bingol B 0	KT.080	Nemrut Dag 1	KT.106	Nemrut Dag 1
KT.015	Bingol B 0	KT.055	Bingol B 0	KT.108	Bingol B 0	KT.083	Bingol A 0	KT.111	Nemrut Dag 1
KT.016	Bingol B 0	KT.056	Bingol B 0	KT.109	Bingol B 0	KT.086	Nemrut Dag 1	KT.119	Nemrut Dag 1
KT.019	Bingol B 0	KT.057	Bingol B 0	KT.112	Bingol B 0	KT.087	Nemrut Dag 1	KT.060	Nemrut Dag 1
KT.023	Bingol B 0	KT.063	Bingol B 0	KT.113	Gollu Dag 0	KT.088	Nemrut Dag 1	RGM-2 1	Nenezi Dag 0
KT.026	Bingol B 0	KT.064	Bingol B 0	KT.115	Bingol B 0	KT.093	Bingol A 0	RGM-2 2	Nenezi Dag 0
KT.027	Bingol B 0	KT.066	Bingol B 0	KT.116	Bingol B 0	KT.096	Nemrut Dag 1	RGM-2 3	Nenezi Dag 0
KT.028	Bingol B 0	KT.067	Bingol B 0	KT.117	Nenezi Dag 0	KT.098	Nemrut Dag 1	RGM-2 4	Nenezi Dag 0
KT.029	Bingol B 0	KT.068	Bingol B 0	KT.001	Bingol A 0	KT.100	Nemrut Dag 1	RGM-2 5	Nenezi Dag 0
KT.030	Bingol B 0	KT.070	Bingol B 0	KT.003	Nemrut Dag 1	KT.103	Nemrut Dag 1	RGM-2 6	Nenezi Dag 0
KT.031	Bingol B 0	KT.072	Bingol B 0	KT.011	Bingol A 0	KT.105	Bingol A 0	RGM-2 7	Nenezi Dag 0
KT.032	Bingol B 0	KT.075	Bingol B 0	KT.012	Nemrut Dag 1	KT.107	Bingol A 0	RGM-2 8	Nenezi Dag 0
KT.033	Bingol B 0	KT.076	Bingol B 0	KT.018	Nemrut Dag 1	KT.110	Nemrut Dag 1	RGM-2 9	Nenezi Dag 0
KT.036	Bingol B 0	KT.077	Bingol B 0	KT.020	Nemrut Dag 1	KT.114	Nemrut Dag 1	RGM-2 10	Nenezi Dag 0
KT.037	Bingol B 0	KT.078	Bingol B 0	KT.021	Bingol A 0	KT.118	Nemrut Dag 1		
		KT.079	Bingol B 0	KT.022	Nemrut Dag 1	KT.120	Bingol A 0		
		KT.081	Bingol B 0	KT.024	Nemrut Dag 1	KT.010	Nemrut Dag 1		
		KT.082	Bingol B 0	KT.025	Nemrut Dag 1	KT.017	Bingol A 0		
		KT.084	Bingol B 0	KT.035	Bingol A 0	KT.034	Bingol A 0		

AP	CA	Sensitivity	Specificity
0.522	0.763	1.000	1.000

Fig. 5.19. Resultados predicción y validación Random Forest

En la Figura 5.20 se observa el resultado del cálculo de las variables más importantes.

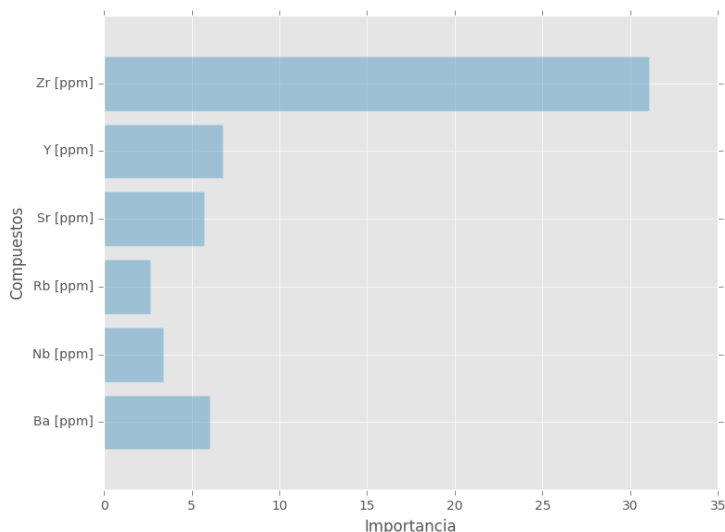


Fig. 5.20. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.20, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Y [ppm] – Zr [ppm] (Figura 5.21) y Y [ppm] – Zr [ppm] – Rb [ppm] (Figura 5.22).

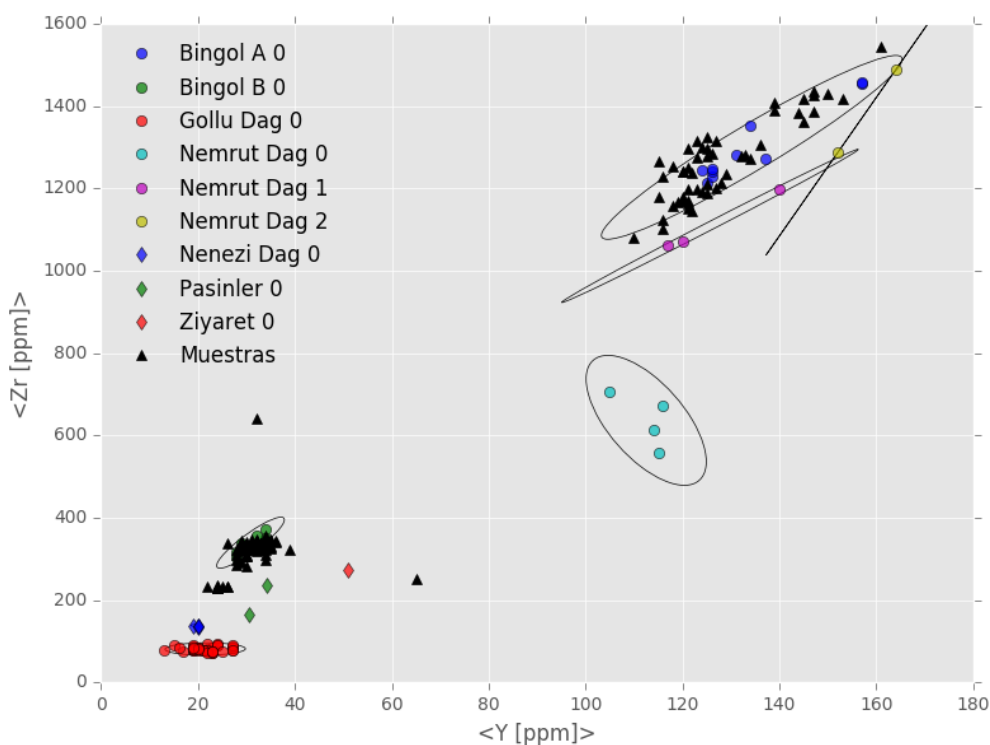


Fig. 5.21. Representación bidimensional Y [ppm] – Zr [ppm]

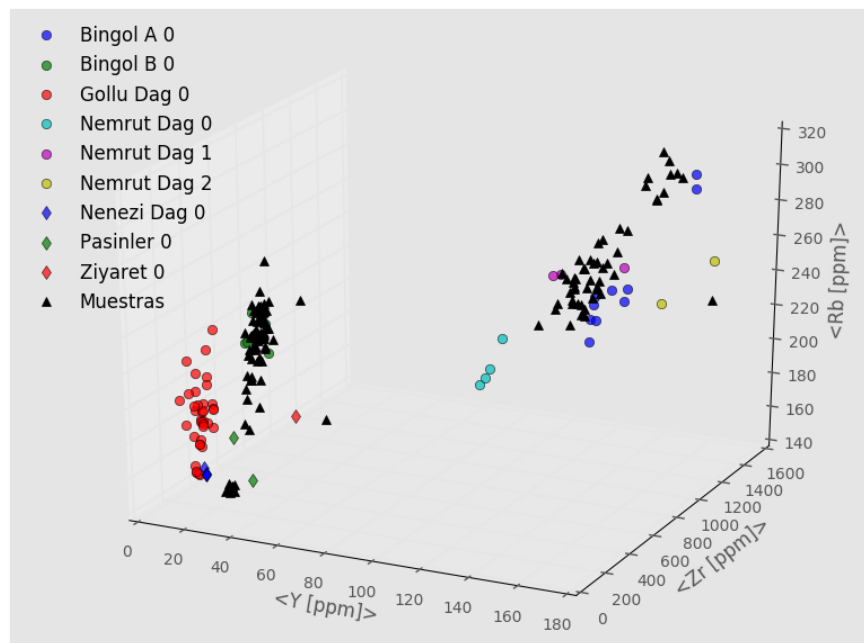


Fig. 5.22. Representación tridimensional Y [ppm] – Zr [ppm] – Rb [ppm]

Tal y como se indica en el artículo en cuestión, las primeras 68 muestras pertenecen al volcán Bingöl B. Hay una discrepancia respecto a las conclusiones del artículo en las muestras KT.046, KT.117 y KT.113. Las dos primeras son asignadas a la fuente Nenezi Dağ 0 y la última a la fuente Göllü Dağ 0. Observando los gráficos anteriores se distingue bien la nube de puntos perteneciente a las muestras de la fuente Bingöl B.

Si nos fijamos en la tabla resumen de la muestra KT.117 podremos observar como los valores de la fuente Nenezi Dağ y Bingöl B no difieren mucho. Si analizamos individualmente dicha muestra seleccionando únicamente estas dos fuentes, la nueva predicción asigna la muestra a la fuente Nenezi Dağ (véase Figura 5.23).

Muestra	Bingol B 0	Nenezi Dag 0	Nulo
KT.117	94	306	0

Fig. 5.23. Tabla resumen Random Forest

Por lo que respecta a las muestras pertenecientes a las fuentes Bingöl A y Nemrut Dağ, existe gran discrepancia entre las conclusiones extraídas en el análisis del programa y las del artículo. Observando los gráficos realizados, se identifica claramente una nube de puntos alrededor de las fuentes Nemrut Dağ y Bingöl A. Sin embargo, la semejanza de estas dos fuentes por lo que respecta a la composición química de determinados elementos, hace difícil la asignación de las muestras, por lo que muchas veces las fuentes asignadas son intercambiadas respecto a las conclusiones del artículo. Además, disponemos de pocos datos de los elementos de las fuentes, cosa que dificulta la comparación con mayor detalle y por tanto, el trabajo de asignación.

A continuación, en la Tabla 5.1., se muestra el resumen de asignación de muestras que no coinciden con los resultados del artículo.

Tabla 5.1. Comparación de resultados del artículo con el programa

Muestra	Procedencia Artículo	Procedencia Source Seeker
KT.046	Bingöl B	Nenezi Dağ 0
KT.113	Bingöl B	Göllü Dağ 0
KT.003	Bingöl A	Nemrut Dağ 1
KT.012	Bingöl A	Nemrut Dağ 1
KT.018	Bingöl A	Nemrut Dağ 1
KT.020	Bingöl A	Nemrut Dağ 1
KT.022	Bingöl A	Nemrut Dağ 1
KT.024	Bingöl A	Nemrut Dağ 1
KT.025	Bingöl A	Nemrut Dağ 1
KT.043	Bingöl A	Nemrut Dağ 1
KT.058	Bingöl A	Nemrut Dağ 1
KT.061	Bingöl A	Nemrut Dağ 1
KT.062	Bingöl A	Nemrut Dağ 1
KT.065	Bingöl A	Nemrut Dağ 1
KT.073	Bingöl A	Nemrut Dağ 1
KT.080	Bingöl A	Nemrut Dağ 1
KT.086	Bingöl A	Nemrut Dağ 1
KT.087	Bingöl A	Nemrut Dağ 1
KT.088	Bingöl A	Nemrut Dağ 1
KT.096	Bingöl A	Nemrut Dağ 1
KT.098	Bingöl A	Nemrut Dağ 1
KT.100	Bingöl A	Nemrut Dağ 1
KT.103	Bingöl A	Nemrut Dağ 1
KT.110	Bingöl A	Nemrut Dağ 1
KT.114	Bingöl A	Nemrut Dağ 1
KT.118	Bingöl A	Nemrut Dağ 1
KT.017	Nemrut Dağ	Bingöl A 0
KT.034	Nemrut Dağ	Bingöl A 0
KT.044	Nemrut Dağ	Bingöl A 0
KT.045	Nemrut Dağ	Bingöl A 0
KT.069	Nemrut Dağ	Bingöl A 0

5.2.6. Artículo 6

Este artículo[22] francés analiza las distintas muestras de obsidiana provenientes del yacimiento de Cafer Höyük, situado a 40 km al nordeste de Malatya, Turquía. En el estudio se realizan representaciones bidimensionales de elementos y diagramas triangulares para identificar las fuentes de origen.

Antes de realizar el análisis de las muestras con el programa, es necesario convertir las unidades de los elementos expresados como porcentaje de oxido a ppm. Para transformar los elementos utilizaremos las ecuaciones Ec.4.2, Ec.4.4, Ec.4.5 y Ec.4.6.

Una vez realizadas las conversiones necesarias se procede con el análisis de procedencia. Para ello se seleccionan todas las muestras y fuentes y se ejecuta el algoritmo. En la Figura 5.24 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
B2	Bingol A 0
B4	Bingol A 0
B12	Bingol A 0
F6	Nemrut Dag 1
F21	Bingol A 0
B8	Bingol B 0
B11	Bingol B 0
F8	Bingol B 0
F10	Bingol B 0
F11	Bingol B 0
F12	Bingol B 0
F20	Bingol B 0

AP	CA	Sensitivity	Specificity
0.688	0.930	1.000	1.000

Fig. 5.24. Resultados predicción y validación Random Forest

En la Figura 5.25 se observa el resultado de las variables más importantes.

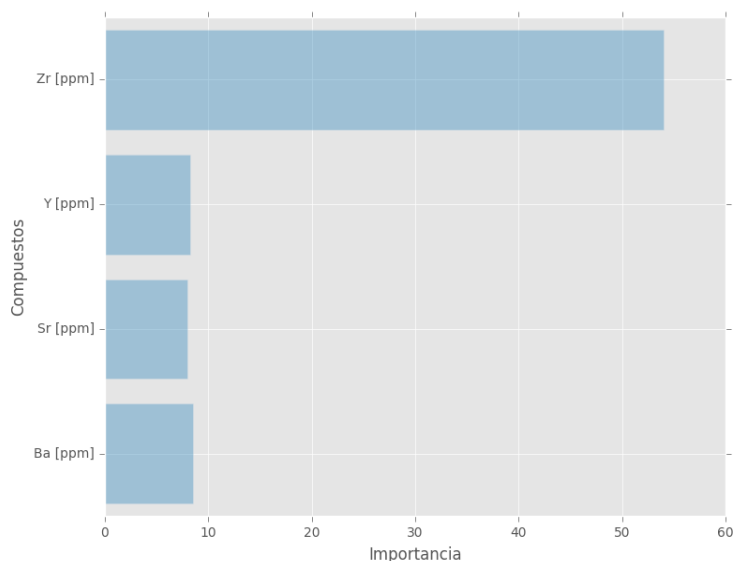


Fig. 5.25. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.25, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Zr [ppm] – Y [ppm] (Figura 5.26) y Y [ppm] – Zr [ppm] – Sr [ppm] (Figura 5.27).

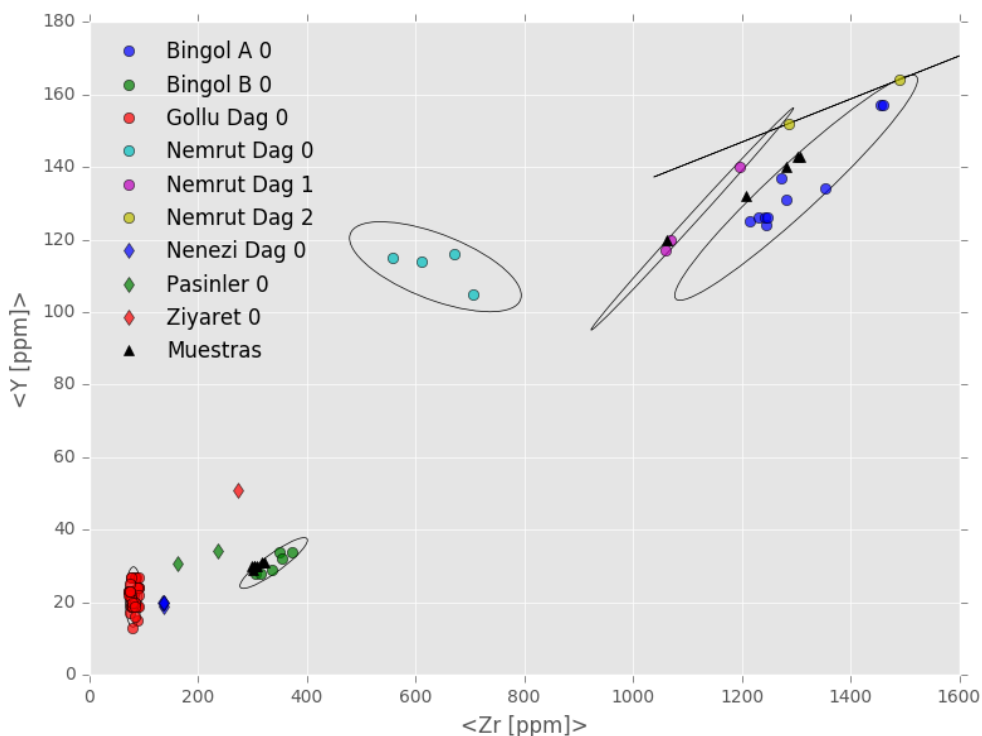


Fig. 5.26. Representación bidimensional Zr [ppm] – Y [ppm]

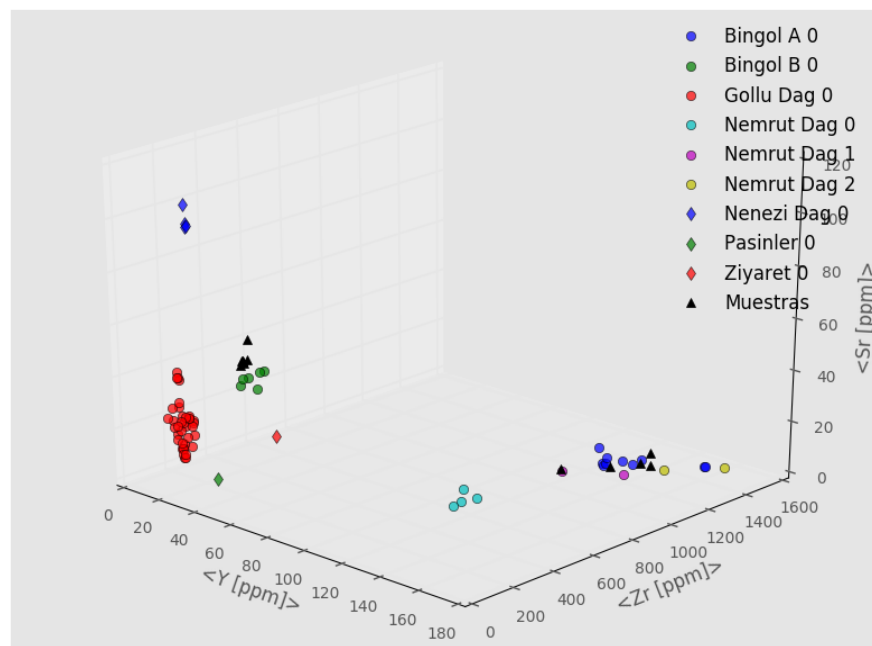


Fig. 5.27. Representación tridimensional Y [ppm] – Zr [ppm] – Sr [ppm]

Observando las representaciones realizadas y la predicción del programa se puede ver como las muestras B2, 4, 12 y F12 pertenecen a la fuente Bingöl A del mismo modo que está indicado en el presente artículo. Sin embargo, la muestra F6 es asignada a la fuente Nemrut Dağ por el programa, mientras que el artículo señala que pertenece a Bingöl A.

Si nos fijamos en la tabla resumen del programa, observamos como existe poca diferencia entre el número de árboles que asignan dicha muestra a Nemrut Dağ y a Bingöl A. Analizando individualmente la muestra obtenemos una nueva tabla, (véase Figura 5.28) y una nueva predicción, asignando la muestra a la fuente Bingöl A. No obstante, tal y como se puede ver en las Figuras 26 y 27, dicha muestra pertenece claramente a la fuente Nemrut Dağ. Del mismo modo que en el análisis anterior, las muestras pertenecientes a Bingöl A y Nemrut Dağ son difíciles de distinguir entre ellas debido a la semejanza de composición química que tienen las diferentes fuentes y el programa no es capaz de asignarlas correctamente.

Muestra	Bingol A 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nulo
F6	303	27	57	13	0

Fig. 5.28. Tabla resumen Random Forest

Por último, las muestras B8 y 11 y F8, 10, 11, 12 y 20 corresponden a la fuente Bingöl B, coincidiendo con las conclusiones del artículo y representadas en los gráficos anteriores.

5.2.7. Artículo 7

El presente artículo[8] estudia muestras de obsidiana procedentes del oeste de Irán, del yacimiento de East Chia Sabz. Para la asignación de las muestras se utilizan representaciones bidimensionales de distintos elementos. Este estudio pone de manifiesto la semejanza de composición química que presentan la obsidiana procedente de Nemrut Dağ con la procedente de Bingöl A. El estudio inicial de las muestras fue realizado mediante fluorescencia de rayos X, aunque fue necesario utilizar el análisis por activación neutrónica para obtener resultados definitivos.

En el presente artículo no es necesario realizar ningún cambio de unidad, por lo tanto se analizan directamente las muestras proporcionadas. Para ello se seleccionan todas las muestras y fuentes disponibles y se ejecuta el programa. En la Figura 5.29 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
HDI001	Bingol A 0
HDI002	Bingol A 0
HDI003	Bingol A 0
HDI004	Bingol A 0
HDI005	Bingol A 0
HDI006	Nemrut Dag 1
HDI007	Bingol A 0
HDI008	Bingol A 0
HDI009	Bingol A 0
HDI010	Bingol A 0
HDI011	Bingol A 0
HDI012	Bingol A 0
HDI013	Bingol A 0
HDI014	Bingol A 0
HDI015	Bingol A 0
HDI016	Bingol A 0
HDI017	Bingol A 0
HDI018	Bingol A 0
HDI019	Bingol A 0
HDI020	Bingol A 0

AP	CA	Sensitivity	Specificity
0.437	0.683	1.000	1.000

Fig. 5.29. Resultados predicción y validación Random Forest

En la Figura 5.30 se observa el resultado del cálculo de las variables más importantes.

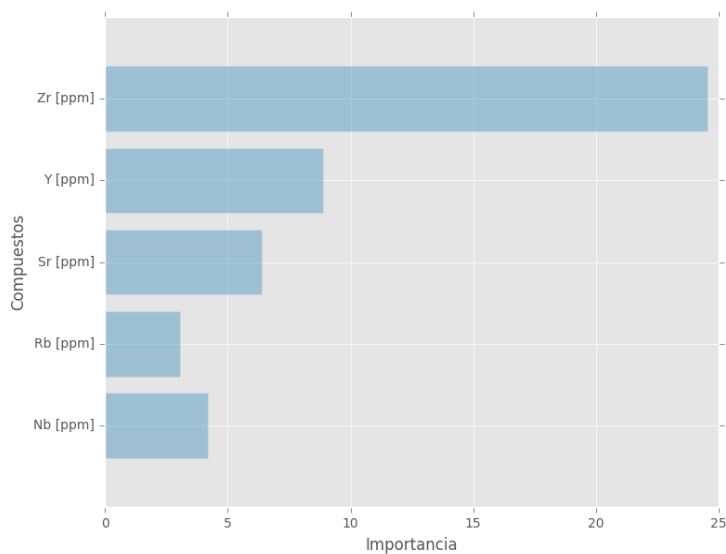


Fig. 5.30. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.30, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Y [ppm] – Zr [ppm] (Figura 5.31) y Y [ppm] – Zr [ppm] – Sr [ppm] (Figura 5.32).

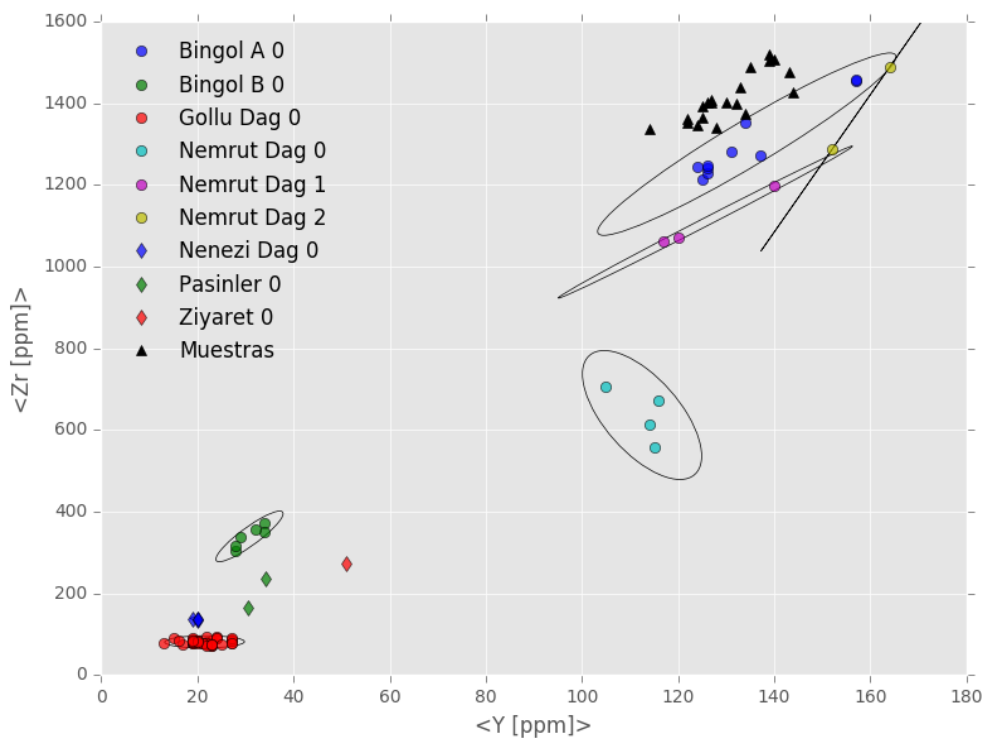


Fig. 5.31. Representación bidimensional Y [ppm] – Zr [ppm]

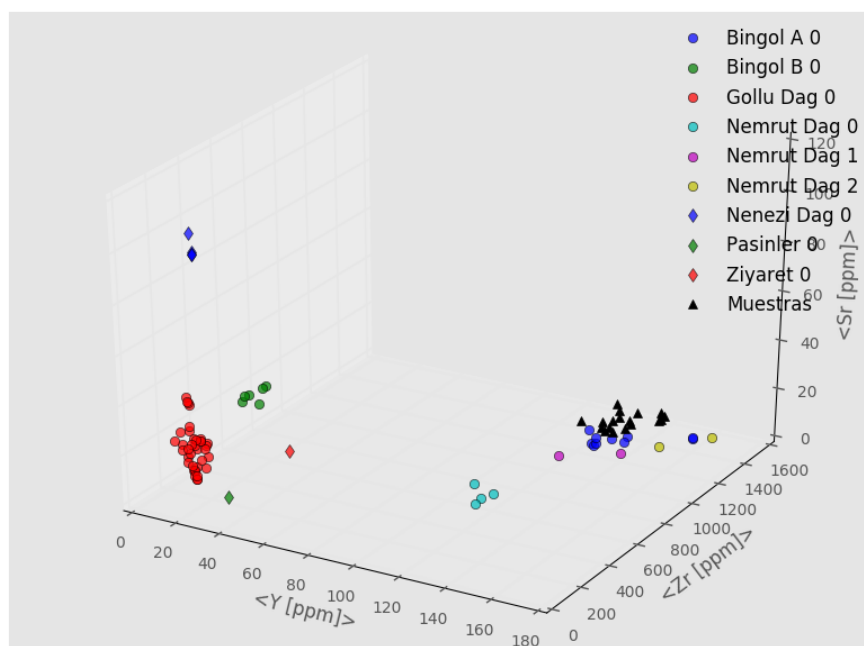


Fig. 5.32. Representación tridimensional Y [ppm] – Zr [ppm] – Sr [ppm]

En este caso, la predicción del Random Forest atribuye todas las muestras a la fuente Bingöl A, excepto la muestra HDI006 que es asignada a la fuente Nemrut Dağ. De acuerdo con las conclusiones del estudio, las muestras son inicialmente asignadas a las fuentes Bingöl A y Nemrut Dağ mediante fluorescencia de rayos X, sin embargo los datos de composición del Br, Cs y Mn analizados mediante el análisis por activación neutrónica determinan que la fuente de origen es Nemrut Dağ.

Observando la tabla resumen del análisis realizado por el programa no se puede asegurar que las muestras pertenezcan a una fuente en concreto. Analizando individualmente cada muestra y centrándonos en las dos fuentes descritas por el artículo tampoco es posible llegar a ninguna conclusión numérica. Ambas fuentes tienen composiciones muy similares y con los datos disponibles el programa no es capaz de distinguir el origen de estas muestras. Sin embargo, utilizando las gráficas representadas anteriormente en las Figuras 5.31 y 5.32 se concluyen que las muestras del presente artículo pertenecen a la fuente Nemrut Dağ.

5.2.8. Artículo 8

El presente artículo[10] introduce la división de la obsidiana según sea alcalina, prealcalina o calcalcalina, y enfoca el estudio a la caracterización de la obsidiana prealcalina, característica de Nemrut Dağ y de la zona de Bingöl A. El estudio revela que las obsidias provenientes de Bingöl A y Nemrut Dağ son distinguibles con las muestras adecuadas de dichas fuentes y un análisis en profundidad de determinados elementos.

Antes de proceder con el análisis de las muestras es necesario convertir los elementos FeO, MnO, CaO y TiO₂ a su forma pura Fe, Mn, Ca y Ti mediante las ecuaciones Ec. 4.1, Ec. 4.3, Ec. 4.5, y Ec. 4.6.

Una vez realizadas las conversiones se ejecuta el algoritmo Random Forest. En la Figura 5.33 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion	EA212	Nemrut Dag 1	EA242	Nemrut Dag 1	EA289	Nemrut Dag 2
EA261	Nemrut Dag 0	EA213	Nemrut Dag 1	EA243	Nemrut Dag 1	EA2810	Nemrut Dag 2
EA262	Nemrut Dag 0	EA214	Nemrut Dag 1	EA244	Nemrut Dag 1	EA2811	Nemrut Dag 2
EA263	Nemrut Dag 0	EA221	Nemrut Dag 1	EA245	Nemrut Dag 1	EA2812	Nemrut Dag 2
EA264	Nemrut Dag 0	EA222	Nemrut Dag 1	EA246	Nemrut Dag 1	EA291	Nemrut Dag 2
EA265	Nemrut Dag 0	EA223	Nemrut Dag 1	EA247	Nemrut Dag 1	EA292	Nemrut Dag 2
EA266	Nemrut Dag 0	EA224	Nemrut Dag 1	EA248	Nemrut Dag 1	EA293	Nemrut Dag 2
EA267	Nemrut Dag 0	EA225	Nemrut Dag 1	EA249	Nemrut Dag 1	EA641	Nemrut Dag 2
EA268	Nemrut Dag 0	EA226	Nemrut Dag 1	EA2410	Nemrut Dag 1	EA642	Nemrut Dag 2
EA269	Nemrut Dag 0	EA227	Nemrut Dag 1	EA2411	Nemrut Dag 1	EA643	Nemrut Dag 2
EA2610	Nemrut Dag 0	EA228	Nemrut Dag 1	EA2412	Nemrut Dag 1	EA644	Nemrut Dag 2
EA2611	Nemrut Dag 0	EA229	Nemrut Dag 1	EA2413	Nemrut Dag 1	EA645	Nemrut Dag 2
EA2612	Nemrut Dag 0	EA2210	Nemrut Dag 1	EA2414	Nemrut Dag 1	EA481	Bingol A 0
EA2613	Nemrut Dag 0	EA2211	Nemrut Dag 1	EA2415	Nemrut Dag 1	EA482	Bingol A 0
EA2614	Nemrut Dag 0	EA2212	Nemrut Dag 1	EA2416	Nemrut Dag 1	EA483	Bingol A 0
EA251	Nemrut Dag 1	EA2213	Nemrut Dag 1	EA2417	Nemrut Dag 1	EA484	Bingol A 0
EA252	Nemrut Dag 1	EA2214	Nemrut Dag 1	EA2418	Nemrut Dag 1	EA485	Bingol A 0
EA253	Nemrut Dag 1	EA2215	Nemrut Dag 1	EA201	Nemrut Dag 2	EA486	Nemrut Dag 2
EA254	Nemrut Dag 1	EA2216	Nemrut Dag 1	EA202	Nemrut Dag 2	EA487	Bingol A 0
EA255	Nemrut Dag 1	EA2217	Nemrut Dag 1	EA203	Nemrut Dag 2	EA488	Nemrut Dag 2
EA256	Nemrut Dag 1	EA2218	Nemrut Dag 1	EA271	Nemrut Dag 2	EA489	Bingol A 0
EA257	Nemrut Dag 1	EA231	Nemrut Dag 1	EA272	Nemrut Dag 2	EA4810	Bingol A 0
EA258	Nemrut Dag 1	EA232	Nemrut Dag 1	EA273	Nemrut Dag 2	EA4811	Bingol A 0
EA259	Nemrut Dag 1	EA233	Nemrut Dag 1	EA281	Nemrut Dag 2	EA471	Bingol A 0
EA2510	Nemrut Dag 1	EA234	Nemrut Dag 1	EA282	Nemrut Dag 2		
EA2511	Nemrut Dag 1	EA235	Nemrut Dag 1	EA283	Nemrut Dag 2		
EA211	Nemrut Dag 1	EA236	Nemrut Dag 1	EA284	Nemrut Dag 2		
		EA237	Nemrut Dag 1	EA285	Nemrut Dag 2		
		EA238	Nemrut Dag 1	EA286	Nemrut Dag 2		
		EA239	Nemrut Dag 1	EA287	Nemrut Dag 2		
		EA241	Nemrut Dag 1	EA288	Nemrut Dag 2		

AP	CA	Sensitivity	Specificity
0.833	0.925	1.000	1.000

Fig. 5.33. Resultados predicción y validación Random Forest

En la Figura 5.34 se observa el resultado del cálculo de las variables más importantes.

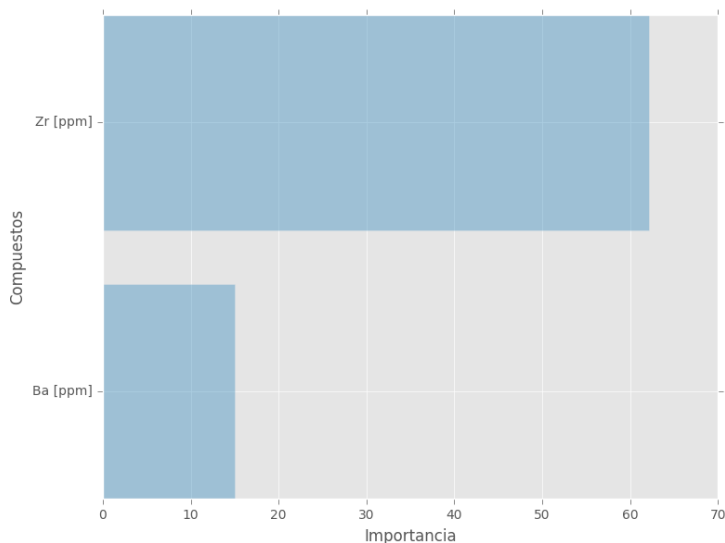


Fig. 5.34. Variables más importantes Random Forest

Debido a que solo se han obtenido dos elementos característicos en la Figura 5.34, se representará el gráfico tridimensional añadiendo otro elemento. A continuación se representan: Ba [ppm] – Zr [ppm] (Figura 5.35) y Zn [ppm] – Zr [ppm] – Ba [ppm] (Figura 5.36).

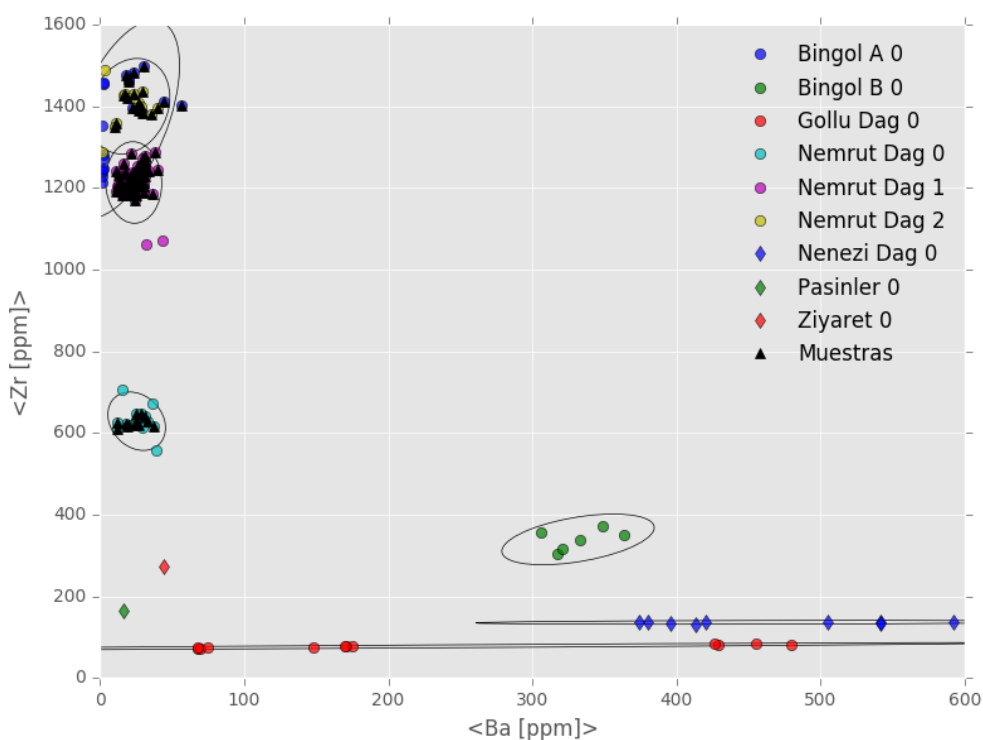


Fig. 5.35. Representación bidimensional Ba [ppm] – Zr [ppm]

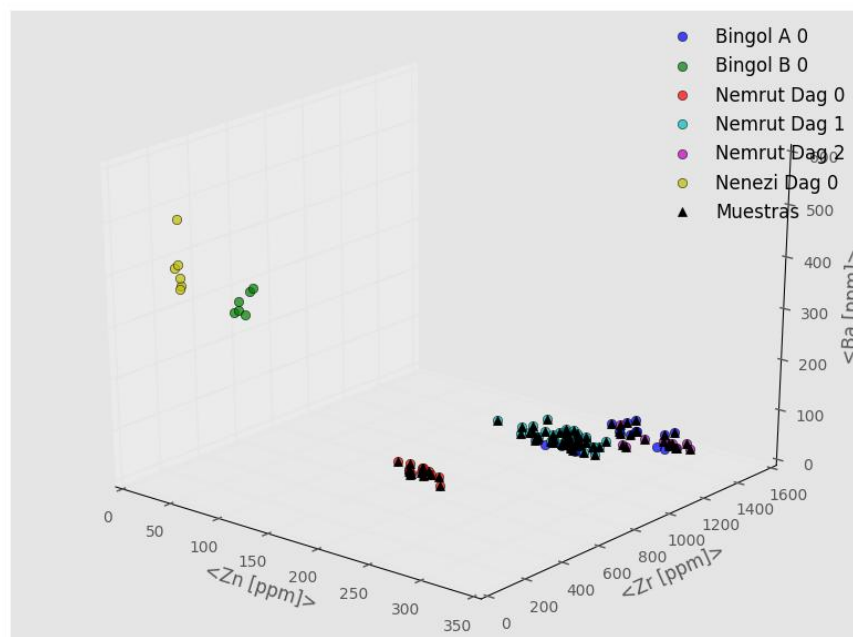


Fig. 5.36. Representación tridimensional Zn [ppm] – Zr [ppm] – Ba [ppm]

Una vez realizado el análisis observamos como todas muestras coinciden exactamente con las fuentes, tanto en la representación bidimensional como en la tridimensional. Este hecho se deba a que el presente artículo ha sido utilizado para crear la base de datos de las fuentes del programa debido a la gran cantidad de muestras que contiene y los abundantes elementos que cada muestra posee.

No obstante, la realización de este análisis permite comprobar el buen funcionamiento del programa y, en particular la robustez del mismo, ya que, al ser las mismas, todas las muestras tienen que coincidir con las fuentes. Por este motivo todos los elementos son asignados correctamente a las fuentes Nemrut Dağ y Bingöl A coincidiendo con la información proporcionada por el artículo.

Sin embargo, las muestras EA486 y EA488 pertenecientes a Bingöl A no son asignadas a dicha fuente. Ya que este artículo ha sido elaborado para crear la base de datos del programa, las conclusiones del análisis tendrían que coincidir con las del artículo. Si nos fijamos en la tabla de resultados, de la muestra EA484 a la EA488 no queda definido el volcán al que pertenecen. Analizando individualmente las muestras se observa como estas son asignadas a la fuente Bingöl A sin excepción. (Véase Figura 5.37).

Muestra	Bingol A 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nulo
EA484	395	1	4	0	0
EA485	393	2	2	3	0
EA486	399	0	1	0	0
EA487	395	1	4	0	0
EA488	391	2	6	1	0

Fig. 5.37. Tabla resumen Random Forest

La semejanza y la dificultad para asignar correctamente muestras pertenecientes a Bingöl A y Nemrut Dağ queda patente una vez más, siendo necesario realizar análisis más exhaustivos.

5.2.9. Artículo 9

En el presente artículo[7] se analizan muestras provenientes de la zona del Éufrates, Siria. Estas muestras son analizadas conjuntamente con muestras de obsidiana provenientes de fuentes situadas en Anatolia.

En el presente artículo no es necesario realizar ningún cambio de unidad, por lo tanto se analizan directamente las muestras proporcionadas. Para ello se seleccionan todas las muestras y fuentes disponibles y se ejecuta el programa. En la Figura 5.38 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion		
CH01	Gollu Dag 0	DJ06	Bingol B 0
CH02	Gollu Dag 0	HL03	Bingol B 0
DJ02	Gollu Dag 0	HL07	Bingol B 0
DJ04	Gollu Dag 0	HL13	Bingol B 0
DJ05	Gollu Dag 0	HL19	Bingol B 0
HL05	Gollu Dag 0	MB06	Bingol B 0
HL06	Gollu Dag 0	HASGrL	Bingol B 0
HL14	Gollu Dag 0	DJ01	Nemrut Dag 1
HL18	Gollu Dag 0	HL04	Nemrut Dag 1
HL21	Gollu Dag 0	HL08	Nemrut Dag 1
JF01	Gollu Dag 0	HL17	Nemrut Dag 1
MB01	Gollu Dag 0	MB10	Nemrut Dag 1
MB02	Gollu Dag 0	TMM	Nemrut Dag 1
MB03	Gollu Dag 0	HL15	Nemrut Dag 1
MB04	Gollu Dag 0	HASGr2a	Nemrut Dag 0
MB05	Gollu Dag 0	HL09	Nemrut Dag 1
MB07	Gollu Dag 0	HL20	Nemrut Dag 1
MB08	Gollu Dag 0	HL22	Nemrut Dag 1
MB09	Gollu Dag 0	HL16	Nenezi Dag 0
DB	Gollu Dag 0	T519	Nenezi Dag 0
HL10	Gollu Dag 0	HL11	Bingol B 0
DJ03	Bingol B 0	HL12	Nemrut Dag 0

AP	CA	Sensitivity	Specificity
0.718	0.909	1.000	1.000

Fig. 5.38. Resultados predicción y validación Random Forest



En la Figura 5.39 se observa el resultado del cálculo de las variables más importantes.

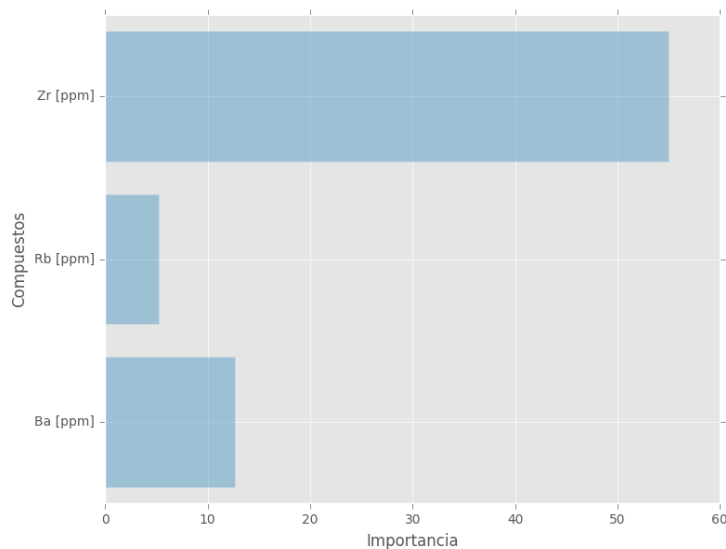


Fig. 5.39. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.39, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Zr [ppm] – Rb [ppm] (Figura 5.40) y Y [ppm] – Zr [ppm] – Sr [ppm] (Figura 5.41).

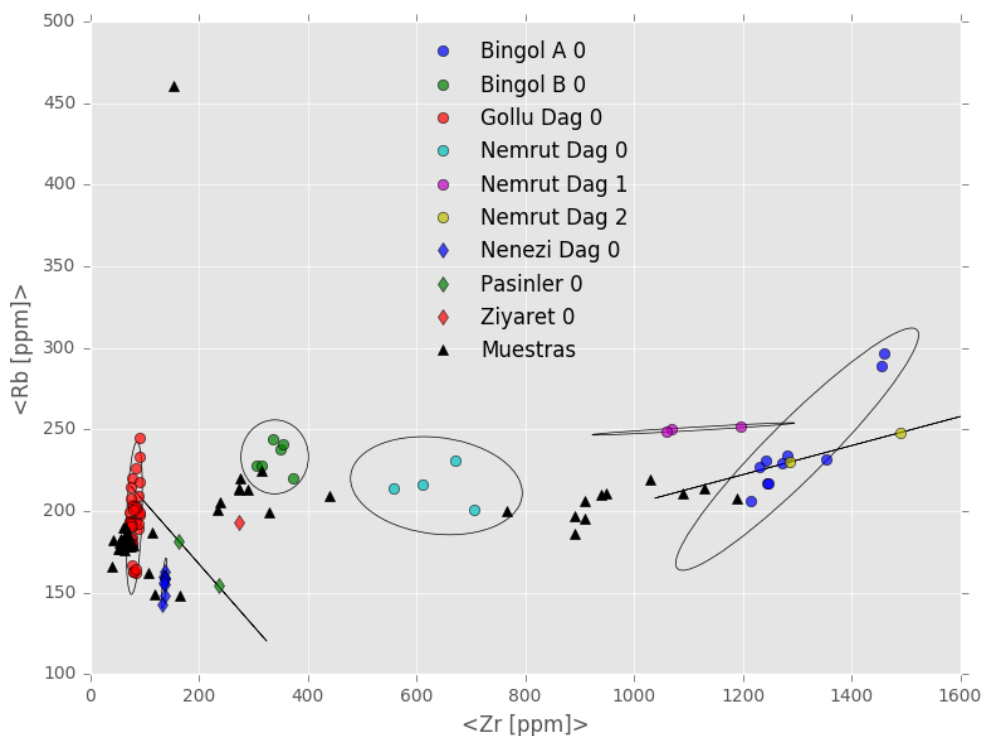


Fig. 5.40. Representación bidimensional Zr [ppm] – Rb [ppm]

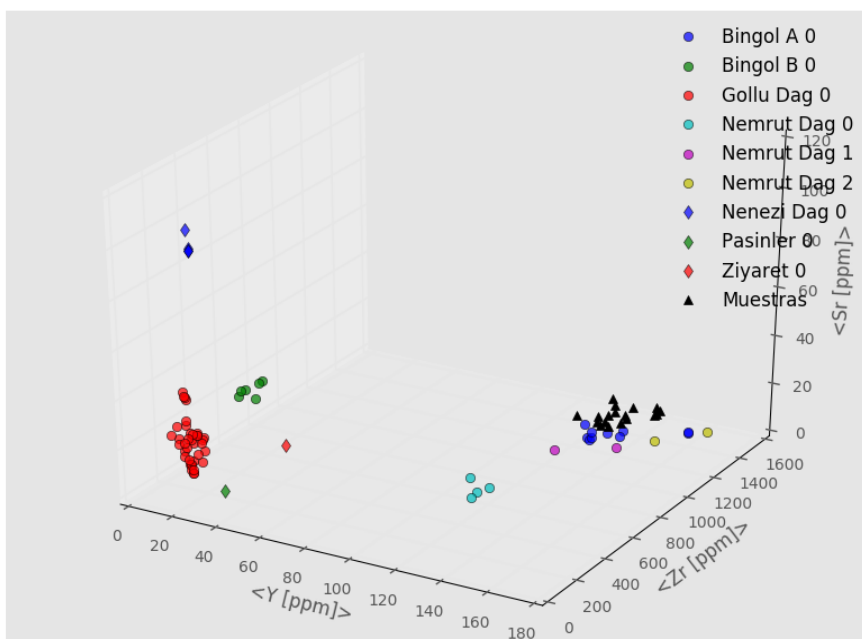


Fig. 5.41. Representación tridimensional Y [ppm] – Zr [ppm] – Sr [ppm]

Analizando los resultados obtenidos y comparando estos con las conclusiones extraídas del artículo se puede observar como las muestras que son asignadas a las fuentes Göllü Dağ, Bingöl B y Nemrut Dağ coinciden con el artículo en cuestión. En la Figura 5.40 se observan las distintas nubes de puntos de las muestras y analizando la tabla resumen del Random Forest se concluye que el programa ha asignado correctamente las muestras,

Por otro lado las muestras HL11 y HL12 son muestras desconocidas en el artículo. Analizando individualmente dichas muestras se concluye que la muestra HL11 pertenece a la fuente Bingöl B tal y como indica la Figura 5.42. Sin embargo la muestra HL12, no pertenece a ninguna fuente conocida por el programa, es la muestra que queda apartada en la Figura 5.40.

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Pasinler 0	Ziyaret 0	Nulo
HL11	7	196	75	54	15	13	14	2	24	0

Fig. 5.42. Tabla resumen Random Forest

5.2.10. Artículo 10

Dicho estudio[19] es similar a los anteriores, se analizan distintas muestras procedentes de la zona del Lago Van, de los yacimientos de Tell Hamoukar y Tell Brak. Se utiliza el método de la representación bidimensional de concentraciones para asignar las distintas muestras. Debido a que es muy difícil distinguir entre la obsidiana de la región de Bingöl A y la de Nemrut Dağ, la atribución a la fuente Bingöl A de las muestras se propone como resultado de la presencia de obsidiana de la fuente Bingöl B.

El presente artículo fue utilizado para elaborar la base de datos del Source Seeker, a continuación se realizará un estudio para comprobar el buen funcionamiento del programa, ya que se comprueba que las fuentes tienen como fuente de origen ellas mismas. Para ello se seleccionan todas las muestras y fuentes y se ejecuta el algoritmo. En la Figura 5.43 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
F2A31	Bingol A 0
F1A1	Bingol B 0
F2B1	Bingol B 0
F3ATW1	Bingol B 0
F12B1	Bingol A 0
F21	Bingol A 0
F7A11	Bingol A 0
F7B21	Bingol A 0
F9B1	Bingol A 0
F9C1	Bingol A 0
F1611	Bingol B 0
F1621	Bingol B 0
BA1	Bingol A 0
BA2	Bingol A 0
BB1	Bingol B 0

AP	CA	Sensitivity	Specificity
0.522	0.763	1.000	1.000

Fig. 5.43 Resultados predicción y validación Random Forest

En la Figura 5.44 se observa el resultado del cálculo de las variables más importantes.

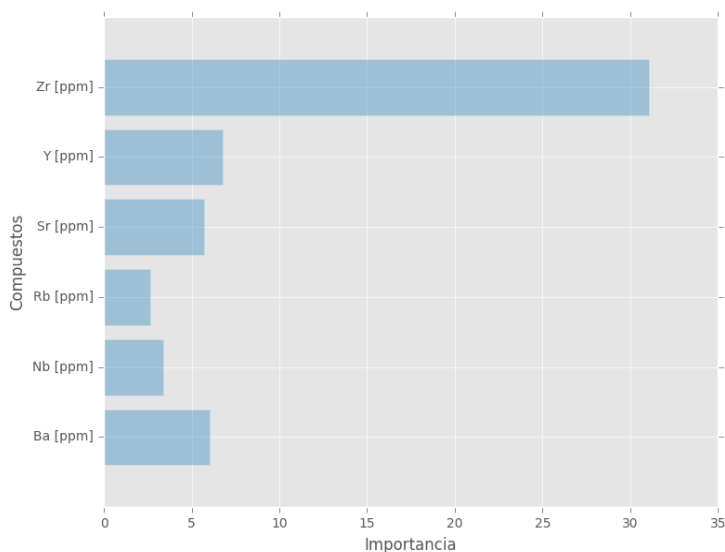


Fig. 5.44. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.44, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Zr [ppm] – Sr [ppm] (Figura 5.45) y Y [ppm] – Zr [ppm] – Sr [ppm] (Figura 5.46).

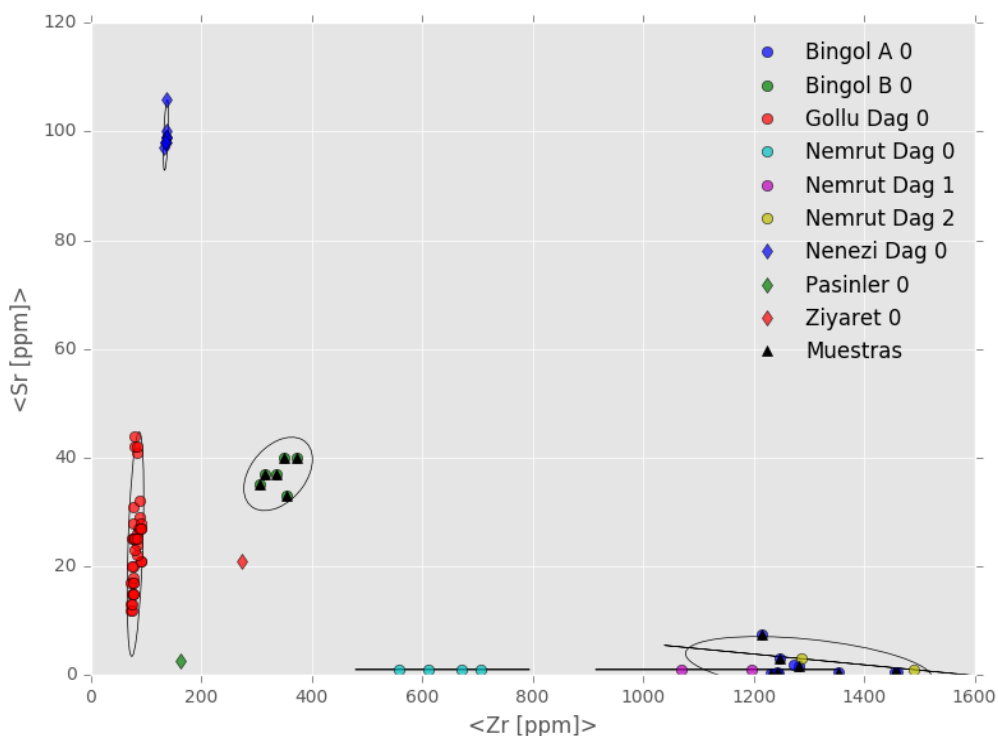


Fig. 5.45. Representación bidimensional Zr [ppm] – Sr [ppm]

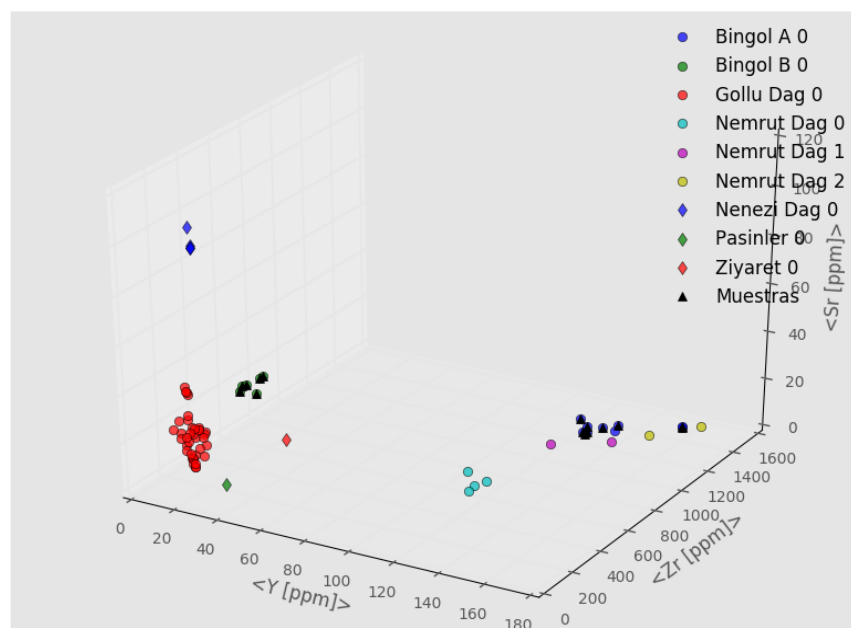


Fig. 5.46. Representación tridimensional Y [ppm] – Zr [ppm] – Sr [ppm]

Las muestras del presente análisis son asignadas con claridad, tanto en los gráficos representados anteriormente como en la tabla resumen del programa se puede observar como cada muestra pertenece a su fuente señalada, Bingöl A y Bingöl B.

En los gráficos anteriores se observa cómo cada muestra coincide con exactitud con su fuente, tal y como se ha explicado anteriormente, esto se debe a que las muestras analizadas son fuentes del programa. Al contrario que en el artículo 8, no existe ninguna problema ni ambigüedad a la hora de adjudicar las muestras entre las fuentes Bingöl A y Bingöl B, esto solo sucede con las fuentes Bingöl A y Nemrut Dağ.

5.3. Artículos nuevos

Una vez finalizado el estudio de procedencia de los diferentes artículos y comprobada la validez y correcta funcionalidad de las mejoras implementadas en el transcurso del presente proyecto se realizó un último estudio de procedencia con dos artículos nuevos proporcionados por el director del proyecto y cuyas muestras eran de origen desconocido.

La finalidad del programa Source Seeker reside en la capacidad de este en asignar e identificar las fuentes de origen de muestras desconocidas. Con el objetivo de comprobar dicha utilidad a continuación se muestra el análisis dos artículos cuyas muestras no han podido ser asignadas a ninguna fuente.

5.3.1. Tell el-Iswid

Este primer artículo[14] analiza una muestra procedente de Tell el-Iswid. La composición química de dicha muestra es comparada con la composición química de las fuentes Nemrut Dağ y Bingöl A. Cabe destacar que dicha comparación no se realiza con diversas muestras de cada fuente sino con un rango de valores para cada elemento de cada fuente.

Con tal de comprobar si nuestro programa es capaz de asignar la muestra a su fuente de origen se realizó un análisis previo seleccionando todas las fuentes disponibles. En la Figura 5.47 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
Messer	Bingol B 0

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Pasinler 0	Ziyaret 0	Nulo
Messer	8	144	17	67	70	87	1	0	6	0

AP	CA	Sensitivity	Specificity
0.718	0.909	1.000	1.000

Fig. 5.47. Resultados predicción y validación Random Forest

En la Figura 5.48 se observa el resultado del cálculo de las variables más importantes.

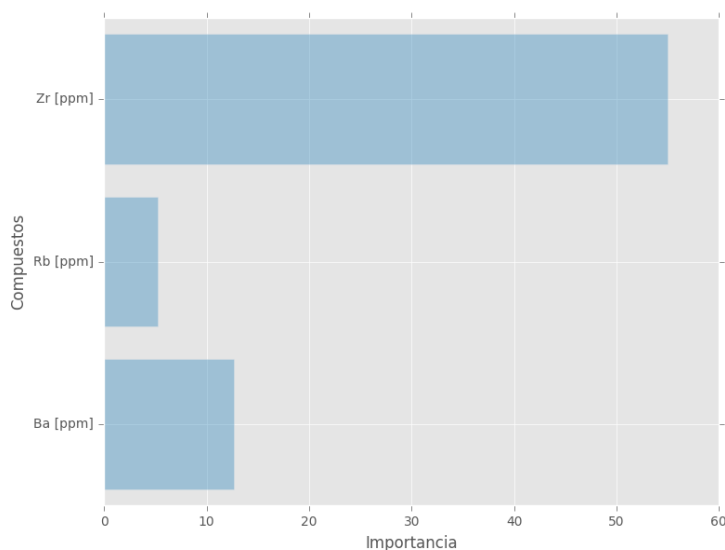


Fig. 5.48. Variable más importante Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.48, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Zr [ppm] – Rb [ppm] (Figura 5.49) y Zr [ppm] – Ba [ppm] – Rb [ppm] (Figura 5.50).

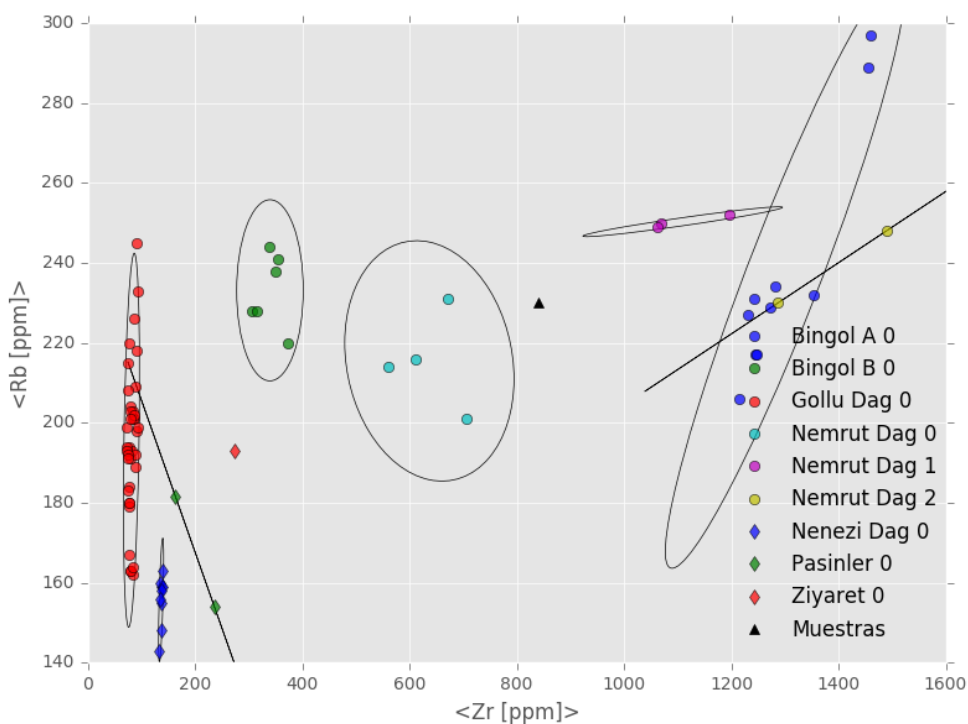


Fig. 5.49. Representación bidimensional Zr [ppm] – Rb [ppm]

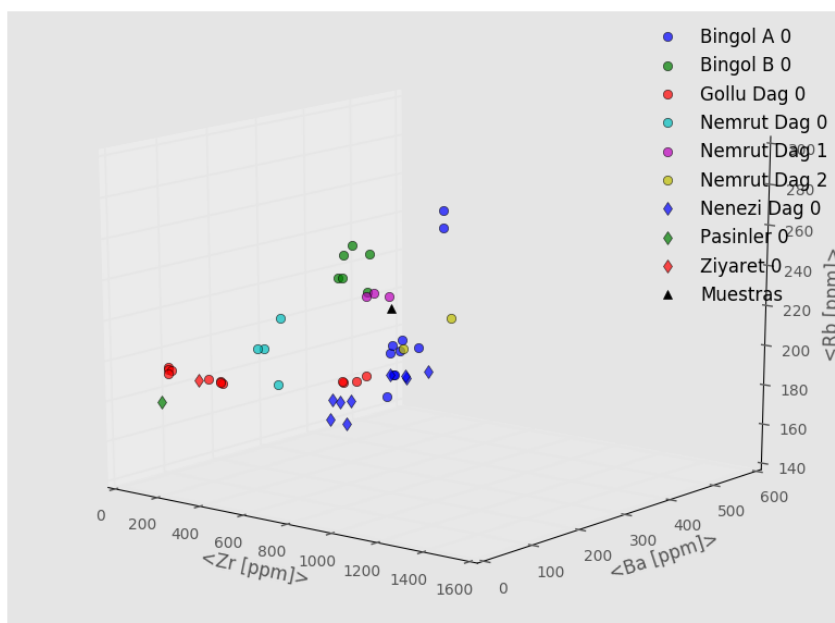


Fig. 5.50. Representación tridimensional Zr [ppm] – Ba [ppm] – Rb [ppm]

Según la predicción del Random Forest, la muestra es asignada a la Fuente Bingöl A, no obstante el número de árboles que asignan esta muestra a dicha fuente es muy pequeño. Estudiando los gráficos representados anteriormente no se puede afirmar que la muestra pertenezca a Bingöl A y en la Figura 5.49 se observa como la muestra está más cerca de la zona de Nemrut Dağ.

Con el fin de realizar un análisis más exhaustivo se realiza un nuevo Random Forest pero esta vez se excluye las fuentes Göllü Dağ, Nenezi Dağ, Pasinler y Ziyaret, ya que en el análisis anterior no fueron relevantes. En la Figura 5.51 se ve la nueva predicción del algoritmo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
Messer	Nemrut Dag 1

Muestra	Bingol A 0	Bingol B 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nulo
Messer	77	28	132	156	7	0

AP	CA	Sensitivity	Specificity
0.892	0.956	1.000	1.000

Fig. 5.51. Resultados predicción y validación Random Forest

La nueva predicción señala que la posible fuente de origen es Nemrut Dağ, para validar la predicción se grafican las diversas muestras en función de los elementos más relevantes indicados por el programa, Ce, Fe y Zr en este caso y se obtiene el gráfico de la Figura 5.52.

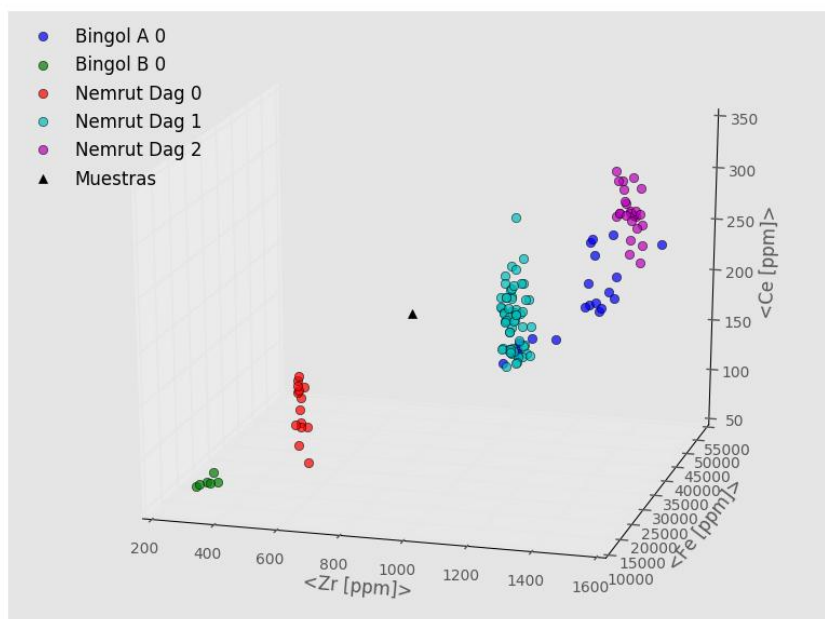


Fig. 5.52. Representación tridimensional Zr [ppm] – Fe [ppm] – Ce [ppm]

Tal y como se observa en la representación anterior no se puede afirmar que la muestra pertenezca a Nemrut Dağ. La base de datos del programa no es muy extensa y existen fuentes que carecen datos de algunos compuestos pudiendo generar errores en el análisis de procedencia, no obstante también es posible que dicha muestra provenga de una fuente desconocida por el programa.

5.3.2. Tell-el-fara

Este segundo artículo[21] estudia una muestra de obsidiana procedente de Tell-el-fara, en concreto un fragmento de un cuchillo. Para ello se realizan diferentes representaciones bidimensionales y se compara la composición química de dicha muestra con la composición química de diferentes fuentes. Finalmente se destaca la fuente Nemrut Dağ como posible fuente de origen.

Para comprobar la hipótesis propuesta en el artículo se realizó un análisis previo con el programa Source Seeker seleccionando todas las fuentes disponibles. En la Figura 5.53 se observan los resultados de la predicción y validación del modelo.

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion
Tell-el-Fara	Nemrut Dag 1

Muestra	Bingol A 0	Bingol B 0	Gollu Dag 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Nenezi Dag 0	Pasinler 0	Ziyaret 0	Nulo
Tell-el-Fara	37	0	9	94	142	14	0	102	2	0

AP	CA	Sensitivity	Specificity
0.676	0.919	1.000	1.000

Fig. 5.53. Resultados predicción y validación Random Forest

En la Figura 5.54 se observa el resultado del cálculo de las variables más importantes.

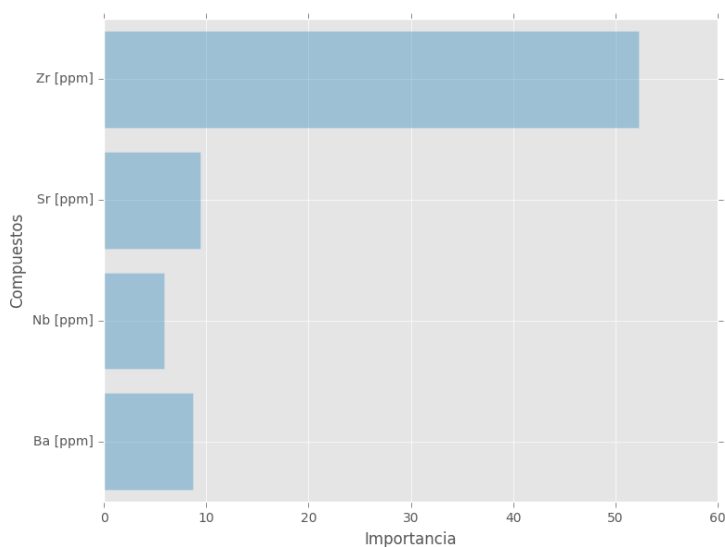


Fig. 5.54. Variables más importantes Random Forest

A partir de los elementos más característicos obtenidos en la Figura 5.54, se representan los gráficos pertinentes, bidimensionales y tridimensionales, que permiten valorar los resultados obtenidos. A continuación se representan los más significativos: Zr [ppm] – Nb [ppm] (Figura 5.55) y Zr [ppm] – Rb [ppm] – Ba [ppm] (Figura 5.56).

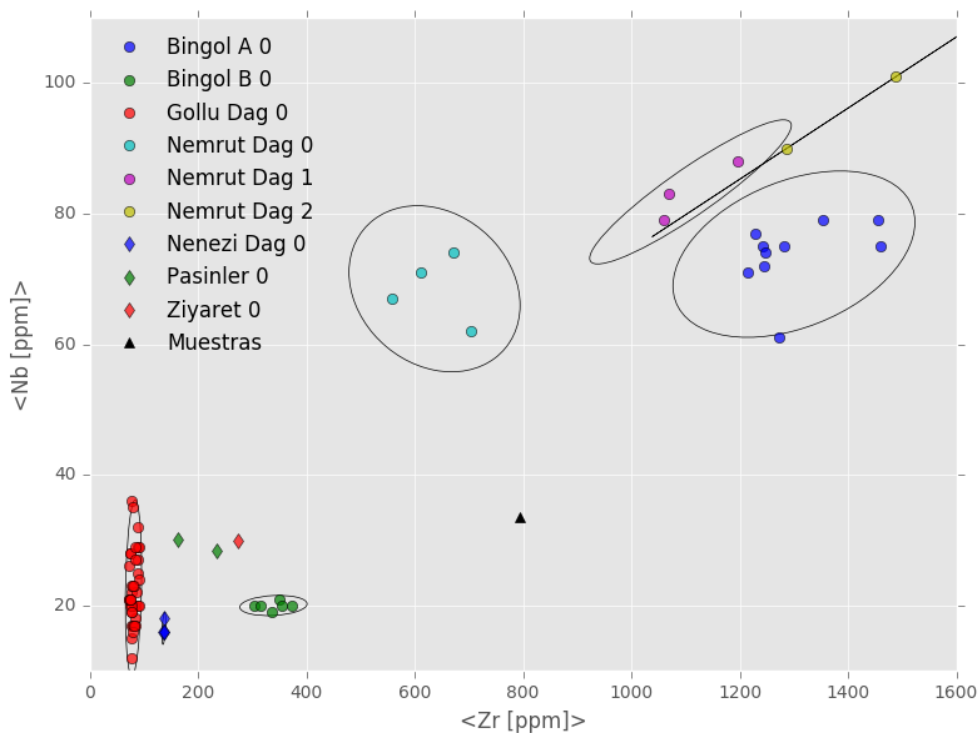


Fig. 5.55. Representación bidimensional Zr [ppm] – Nb [ppm]

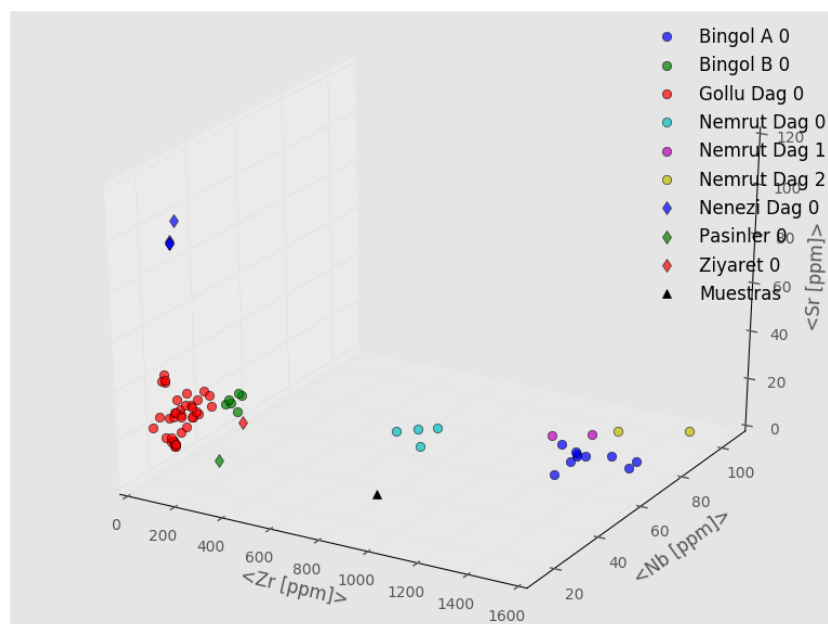


Fig. 5.56. Representación tridimensional Zr [ppm] – Nb [ppm] – Sr [ppm]

La predicción del programa señala a la fuente Nemrut Dağ como posible fuente de origen. No obstante, no difiere mucho de Pasinler y en los gráficos representados anteriormente no se observa que pertenezca a ninguna fuente en concreto. Por este motivo se realiza un nuevo análisis con las fuentes Bingöl A, Nemrut Dağ y Pasinler como únicas fuentes posibles. (Véase Figura 5.57)

- Numero de arboles: 400
- Splitter: UnknownsAsSelector
- Descender: UnknownMergeAsSelector
- Profundidad: 99
- Eliminar compuestos que no contiene la muestra : True
- Eliminar compuestos que no explican todas las fuentes : True

Muestra	Prediccion					
Tell-el-Fara	Nemrut Dag 0					

Muestra	Bingol A 0	Nemrut Dag 0	Nemrut Dag 1	Nemrut Dag 2	Pasinler 0	Nulo
Tell-el-Fara	47	213	56	26	58	0

AP	CA	Sensitivity	Specificity
0.736	0.857	-	1.000

Fig. 5.57. Resultados predicción y validación Random Forest

Este nuevo análisis señala la fuente Nemrut Dağ como fuente de origen, ahora con más claridad que antes, para comprobar su validez se representan los nuevos compuestos señalados por el programa como compuestos más relevantes, Zr y Ce. (Véase Figura 5.58).

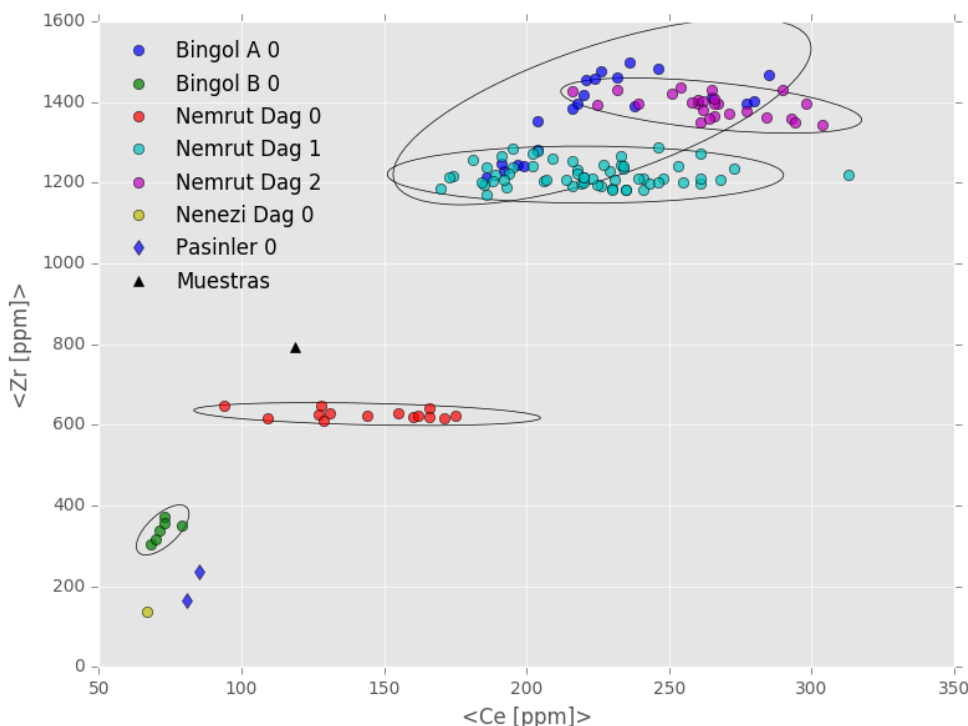


Fig. 5.58. Representación bidimensional Ce [ppm] – Zr [ppm]

Una vez más, no se puede asegurar que dicha muestra pertenezca a Nemrut Dağ, aún así, esta es la más probable, tal y como se observa en la Figura 5.58, la fuente Nemrut Dağ es la fuente más próxima a la muestra, por este motivo las conclusiones de este análisis coinciden con las extraídas del artículo.

Estos resultados coincidirían con el análisis de otras muestras de obsidiana encontradas en Egipto [12]. Estos estudios, llevados a cabo con el programa Source Seeker, mostraron que la obsidiana egipcia no procede de Oriente Próximo sino de volcanes que estarían situados en el Valle del Rift etíope, por tanto, en una siguiente etapa, se debería estudiar la procedencia de estas dos muestras incluyendo en la base de datos del programa valores de composición química de obsidianas extraídas de volcanes etíopes.

6. Vías de continuación

Del mismo modo que los proyectos anteriores, el presente proyecto está orientado a seguir siendo desarrollado en un futuro, por ese motivo a continuación se muestran algunas de las opciones y herramientas que, conjuntamente con el director del proyecto, han sido propuestas para orientar el desarrollo y la continuación de este programa.

6.1. Soporte para otras extensiones

Esta primera medida ya fue propuesta por el anterior proyectista[1] pero a la hora de realizar este proyecto se priorizaron otras mejoras. El programa sólo tiene soporte para ficheros Excel 97-2013, es decir, ficheros con la extensión xls, por lo que la extensiónxlsx de las nuevas versiones de Excel no es compatible. Dicha limitación es debida a los módulos de escritura (xlwt) y lectura (xlrd) que son utilizados para el desarrollo del programa.

La incompatibilidad con las nuevas versiones de Excel no supone ningún problema, ya que siempre se puede guardar el fichero con la versión antigua y utilizarlo para el análisis. Sin embargo, resultaría mucho más cómodo que el programa fuera compatible con todas las versiones existentes. Para realizar la mejora sería necesario utilizar otros paquetes de Python de lectura y escritura de Excel que fuesen compatibles con todas las extensiones existentes.

6.2. Ampliación de las fuentes existentes

Actualmente, la base de datos del programa contiene pocas muestras de cada fuente. Por ejemplo las fuentes Ziyaret y Pasinler solo contienen una muestra cada una, siendo Nemrut Dağ la fuente de la cual se poseen más muestras.

Para poder realizar estudios de procedencia detallados y analizar la existencia de posibles subgrupos en las fuentes es necesario disponer de un gran número de muestras. Aumentar la base de datos permitiría al programa realizar predicciones más precisas. Para ello, del mismo modo que se ha hecho para crear la base de datos inicial, se pueden extraer muestras de las fuentes de artículos como los estudiados a lo largo del presente proyecto.

6.3. Registro de nuevas fuentes

Tal y como se ha podido observar a lo largo del presente proyecto, el programa dispone de siete fuentes de origen distintas, Bingöl A, Bingöl B, Göllü Dağ, Nemrut Dağ, Nenezi Dağ, Pasinler y Ziyaret. No obstante, existen más fuentes de origen de obsidiana en Oriente Próximo, como por ejemplo Meydan Dağ, Suphan Dağ, Gutansar, etc. Por otro lado, como se ha dicho anteriormente, la base de datos del programa debería ampliarse para incluir muestras de otras zonas geográficas como Etiopía, Kenia o los volcanes de Yemen.

Disponer de una gran variedad de fuentes de origen permite asignar muchas más muestras, de origen desconocido hasta el momento. En los análisis del apartado anterior no se pudieron asignar ambas muestras ya que posiblemente pertenecían a otra fuente. Para disponer de un programa mucho más completo, aumentar las fuentes de obsidiana sería una mejora muy interesante a realizar.

6.4. Elipsoide de error

El programa actualmente sólo contempla el cálculo de elipses de error bidimensionales. Introducir el cálculo de elipses de error en el espacio tridimensional fue una mejora propuesta por el anterior proyectista la cuál sería interesante implementar[1]. De este modo, se podría graficar un tercer compuesto y obtener mucha más información. (Véase Figura 6.1)

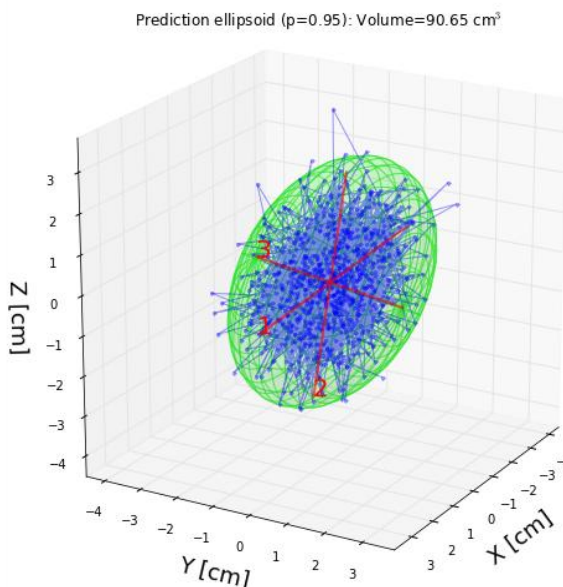


Fig. 6.1. Ejemplo de aplicación elipsoide [9]

Para una primera aproximación a la creación de la elipse de error tridimensional consultar la página web. [9]

6.5. Exportar resultados

Debido a que no es necesario si el número de muestras no es muy elevado, sería interesante poder exportar la tabla que proporciona el Random Forest donde se hallan las muestras y su predicción final a un archivo Excel. Con esta implementación también se podría extraer las tablas resumen creadas en el presente trabajo, para una posible representación gráfica o tratamientos de datos posteriores.

6.6. Creación de una página web

Con el fin de mejorar la divulgación y su facilidad de uso una de las mejoras a implementar en el próximo proyecto sería la creación de una página web que contenga las mismas herramientas que el programa Source Seeker. De esta manera, la página web funcionaría como programa y se podría acceder a ella desde cualquier ordenador y sin necesidad de conocimientos informáticos.

7. Costes

El desarrollo de un proyecto siempre conlleva un coste económico, tanto de personal como de material. En primer lugar, los costes de la mano de obra vienen dados por las horas trabajadas por la persona o personas que han realizado el proyecto. El segundo incluye todos los gastos relacionados con la compra de material, maquinaria utilizada y coste energético que consumen.

Este proyecto no requiere una gran inversión de material para ser llevado a cabo, al ser un programa informático los gastos son más reducidos que otros tipos de proyectos y mucho menos significativos que la mano de obra.

A continuación se muestra en la Tabla 7.1 el desglose de los costes asociados a la realización del presente proyecto.

Tabla 7.1. Costes asociados a la realización del proyecto

MANO DE OBRA				
Concepto	Coste hora	Horas	Coste total	
Mano de obra	25 €/h	550 h	13750 €	
COSTES ASOCIADOS				
Concepto	Coste total	Vida útil	Uso	Imputable
Impresión	12 €	-	-	12 €
Encuadernación	10 €	-	-	10 €
Amortización impresora	400 €	6 años	7 meses	38,89 €
Amortización ordenador	600 €	6 años	7 meses	58,34 €
Costes eléctricos				
Concepto	Consumo	Coste hora	Horas	Coste total
Costes eléctricos	250 W	0,15 €/kWh	550 h	20,63 €
			TOTAL	13889,86 €

Los costes totales ascienden a **13889,86 €**.

8. Planificación

A continuación se muestra la planificación llevada a cabo para realizar el presente proyecto desglosada en la Tabla 8.1 y representada en el siguiente diagrama de Gantt (Figura 8.1).

Tabla 8.1. Planificación del proyecto

Actividad	Inicio	Finalización	Duración
Trabajo Fin de Grado (TFG)	13/09/2016	20/04/2017	219
Presentación del proyecto	13/09/2016	13/09/2016	1
Lectura de los proyectos anteriores	13/09/2016	04/10/2016	21
Contextualización	26/09/2016	04/10/2016	8
Lectura de los artículos	04/10/2016	25/10/2016	21
Excel Primeros 5 artículos	04/10/2016	18/10/2016	14
Primera reunión proyectista anterior	25/10/2016	25/10/2016	1
Familiarización con el programa	25/10/2016	04/11/2016	10
Análisis previo	02/11/2016	18/11/2016	16
Selección de las mejoras	18/11/2016	21/11/2016	3
Implementación de las mejoras	21/11/2016	19/12/2016	28
Excel siguientes 5 artículos	28/11/2016	16/12/2016	18
Segunda reunión proyectista anterior	13/12/2016	13/12/2016	1
Información ejecutables	13/12/2016	22/12/2016	9
Realización del ejecutable	19/12/2016	20/01/2017	32
Análisis de los artículos	19/11/2016	04/12/2016	15
Asignación de muestras desconocidas	29/11/2016	13/12/2016	14
Documentación del trabajo	15/01/2017	20/04/2017	95
Redactado de la memoria	15/01/2017	20/02/2017	36
Revisiones de la memoria	03/03/2017	09/03/2017	6
Preparación de la presentación	26/04/2017	07/05/2017	11
Defensa del trabajo	08/05/2017	18/05/2017	1

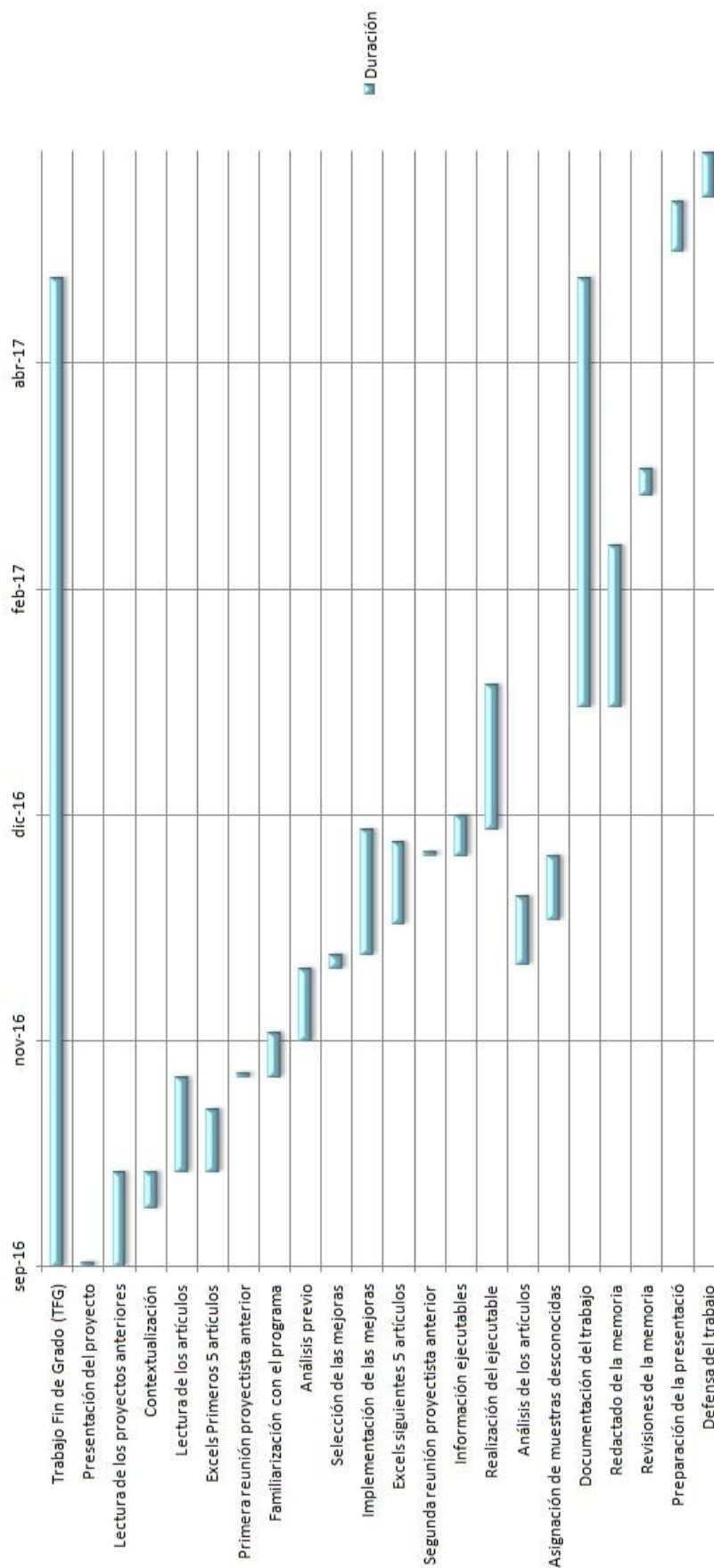


Fig. 8.1. Planificación del proyecto representada en un diagrama de Gantt

9. Impacto ambiental

Dado las características del presente trabajo, el impacto ambiental que genera su implementación es prácticamente menospreciable. Al no contar con una fase experimental, el impacto ambiental se reduce a las emisiones de CO₂ derivadas del consumo eléctrico necesario para su realización y los residuos generados por el uso de consumibles.

Para calcular el consumo asociado a la realización del trabajo, teniendo en cuenta el consumo del ordenador y las horas dedicadas (véase Costes), se multiplican dichos valores y obtenemos un consumo total de unos 137,5kWh. Según la Comisión Europea, 1kWh produce 0,65kg de CO₂, por lo tanto el consumo total representa 89,375kg de CO₂.

Por otro lado, los consumibles necesarios para la redacción del proyecto son papel en formato A4, tinta de impresora y diverso material de oficina.

El consumo de papel se estima en 100 hojas en formato A4. Una hoja A4 mide 297x210mm por lo que el total representa unos 6,24m² de papel. Con una densidad de 80g/m², el peso total utilizado es 0,50kg.

Conclusiones

El presente trabajo cumple con la gran mayoría de los objetivos planteados inicialmente. No obstante, aún existen aspectos del programa potencialmente mejorables tal y como se ha propuesto en el apartado vías de continuación.

La implementación de la tabla resumen del Random Forest permite obtener más información sobre los resultados del estudio realizado por el programa. Gracias a ella es posible validar la predicción según el número de árboles de decisión que hayan asignado la muestra a una fuente en concreto. Esto permite descartar ciertas fuentes y centrar el estudio en las potenciales fuentes de origen.

Las mejoras introducidas en la selección de variables permiten deseleccionar y seleccionar las muestras con mayor facilidad. Dicha mejora es realmente útil en el caso de realizar un estudio con un número de muestras elevado. Combinando esta herramienta con la tabla resumen se pueden realizar estudios más precisos y fiables.

No obstante, en la realización de los estudio de procedencia de los diferentes artículos ha quedado patente la dificultad de asignación de las muestras cuando estas pueden pertenecer a Nemrut Dağ o Bingöl A. La semejanza en la composición que comparten estas fuentes junto con la escasez de muestras que se disponen de la fuente Bingöl A, dificultan la distinción entre ambas fuentes de origen. Además, la falta disponibilidad de un gran número de fuentes de origen y una amplia base de datos ha dificultado la asignación de las nuevas muestras de origen desconocido.

Como muestra el estudio, el algoritmo Clustering K-means es una herramienta útil y eficaz para la caracterización de las fuentes de origen y la elaboración de la base de datos. Esta herramienta permite calcular los diferentes subgrupos internos que contienen las nuevas fuentes introducidas en la base de datos y crea los distintos conglomerados.

La realización de un ejecutable resultó ser el principal problema que presentó el trabajo en su desarrollo. Inicialmente fueron seleccionados tres posibles programas para implementar esta mejora. Sin embargo, la existencia de bugs en algunos de los programas y la utilización de un gran número de librerías como por ejemplo la librería Orange impidieron llevar a cabo esta mejora de la manera deseada. No obstante la utilización de archivos batch para la creación del ejecutable y su instalador resolvieron con eficacia el problema planteado.

Agradecimientos

En primer lugar me gustaría agradecer al director del proyecto el profesor Francisco Javier Giménez su dedicación, soporte, orientación y supervisión del mismo. También quisiera agradecer a Lluís Solano, codirector del proyecto, las sugerencias y toda la ayuda prestada con la realización del ejecutable.

Quisiera agradecer también a Joel Moriana, por la orientación, la explicación del funcionamiento del programa y los análisis de procedencia y sobre todo por la ayuda prestada en la realización de las mejoras.

Finalmente, agradecer a Gloria Rodríguez Varela, Alberto Mateos Bartolomé y Ànnia Mir Garrés, por su desinteresada ayuda aportando información y ayuda en la realización de dicho proyecto.

Bibliografía

Referencias bibliográficas

- [1] Moriana Becerra, Joel. *Desarrollo de software estadístico para el análisis de material arqueológico de Oriente Próximo y Medio*. Barcelona: Escola Tècnica Superior d'Enginyeria Industrial de Barcelona, 2015.
- [2] Binder D, Gratuze B, Mouralis D, Balkan-Atlı N. *New investigations of the Göllüdağ obsidian lava flows system: a multi-disciplinary approach*. J Archaeol Sci. 2011; 3174-3184.
- [3] Boulanger M, Davis R, Glascock M. *Preliminary characterization and regional comparison of the Dasht-i-Nawue obsidian source near Ghazni, Afghanistan*. J Archaeol Sci. 2012; 2320-2328.
- [4] Bressy C, Poupeau G, Yener K.A. *Cultural interactions during the Ubaid and Halaf periods: Tell Kurdu (Amuq Valley, Turkey) obsidian sourcing*. J Archaeol Sci. 2005; 1560-1565.
- [5] Carter T, Grant S, Kartal M, Coşkun A, Özkaya V. *Networks and Neolithisation: sourcing obsidian from Körtik Tepe (SE Anatolia)*. J Archaeol Sci. 2013; 556-569.
- [6] Carter T, Shackley MS. *Sourcing Obsidian From Neolithic Çatalhöyük (Turkey) Using Energy Dispersive X-Ray Fluorescence*. Archaeometry. 2007; 437-454.
- [7] Cauvin M.C, Keller J, Pernicka E. *Obsidian from Anatolian sources in the Neolithic of the Middle Euphrates region (Syria)*. In: Paléorient. 1997, Vol. 23 N°1.pp. 113-122.
- [8] Darabi H, Glascock M. *The source of obsidian artefacts found at East Chia Sabz, Western Iran*. J Archaeol Sci. 2013; 3804-3809.
- [9] Duarte Marcos. *Prediction ellipse and prediction ellipsoid*. Jupyter nbviewer. Consultado 25/01/2017
<http://nbviewer.jupyter.org/github/demotu/BMC/blob/master/notebooks/PredictionEllipseEllipsoid.ipynb>
- [10] Frahm E. *Distinguishing Nemrut Dağ and Bingöl A obsidians: geochemical and landscape differences and the archaeological implications*. J Archaeol Sci. 2012; 1436-1444.
- [11] Giménez J. *Egyptological Information from Chemical Analyses: The Provenance of Obsidian and Glass*. DAMQĀTUM. 2014, N°10; 03-10.
- [12] Giménez J, Sánchez J.A, Solano LL. *The journal of Egyptian Archaeology*. The Egypt Exploration Society. 2015, Vol. 101; 349-359.

- [13] Gratuze B. *Obsidian Characterization by Laser Ablation ICP-MS and its Application to Prehistoric Trade in the Mediterranean and the Near East: Sources and Distribution of Obsidian within the Aegean and Anatolia*. J Archaeol Sci. 1999; 26: 869-881.
- [14] Hendrickx S, Friedman R.F, Cialowicz K.M, Chlodnicki M. *A bladelet core from Tell El-Fara'In*. Orientalia Lovaniensia Analecta. 2004; 606-619
- [15] Modern Henry. *Convertir un programa Bat a Exe, y ponerle icono*. Consultado 18/01/2017 <https://www.youtube.com/watch?v=wCSd2rFq5cQ>
- [16] Modern Henry. *Crear un programa con menus con batch*. Consultado 16/01/2017 <https://www.youtube.com/watch?v=nUEU4WVEXew>
- [17] Howes, Mauricio Andrés Alva. *Implementación y aplicación de herramientas para los estudios de procedencia de objetos arqueológicos de obsidiana*. Barcelona: Escola Tècnica Superior d'Enginyeria Industrial de Barcelona, 2014.
- [18] Keller J, Seifried C. *The Present Status of Obsidian Source Identification in Anatolia and the Near East*. Mineralogisch-Petrographisches Institut: Universitat Freibrug. 57-87.
- [19] Khalidi L, Gratuze B, Boucetta S. *Provenance of Obsidian Excavated From Late Chalcolithic Levels At the Sites of Tell Hamoukar and Tell Brak, Syria*. Archaeometry. 2009; 879-893.
- [20] Milić M. *PXRF characterization of obsidian from central Anatolia, the Aegean and central Europe*. J Archaeol Sci. 2014; 285-296.
- [21] Pernicka E. *Analyse eines prädynastischen Obsidianmessers aus Uneterägypten*. 286-287.
- [22] Saroglu F, Besnus Y, Balkan N, Cauvin M.C. *Origine de l'obsidienne de Cafer Höyük (Turquie): premiers résultats*. In: Paléorient. 1986, Vol. 12 N°2.pp.89-97.
- [23] WinRAR, Consultado 20/01/2017 <https://www.winrar.es/?PHPSESSID=23861c126c35120d3cc72adfad6a44c9>

ANEXO I

Manual de instalación

A continuación se presenta el manual enfocado al usuario, con el fin de facilitar la instalación del programa en Windows y el procedimiento a seguir para su correcta utilización.

El primer paso es extraer el archivo SourceSeeker2.1.rar en la carpeta deseada (véase Figura 1). El archivo contiene: el instalador de Python, el instalador de paquetes, el programa Source Seeker y el ejecutable del programa.

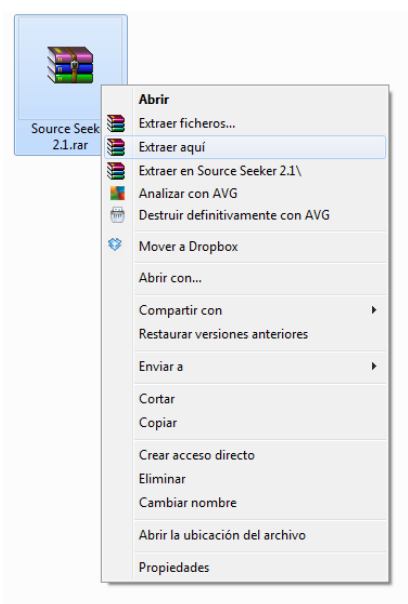


Fig. 1. Archivo SourceSeeker2.1.rar

El siguiente paso a seguir es instalar Python. Para ello, es necesario hacer doble clic en el archivo "python-2.7.13.msi y seguir los pasos indicados.

Una vez Python esté instalado es necesario introducir las variables de entorno, ya que por defecto no se instalan. Para introducir dichas variables es necesario dirigirse a la carpeta de *Equipo* y seleccionar las *Propiedades del sistema*. Una vez allí clicamos *Configuración avanzada del sistema*, esta orden abrirá una ventana pequeña en la cual se encuentra la pestaña de las variables de entorno, tal y como indica la Figura 2.

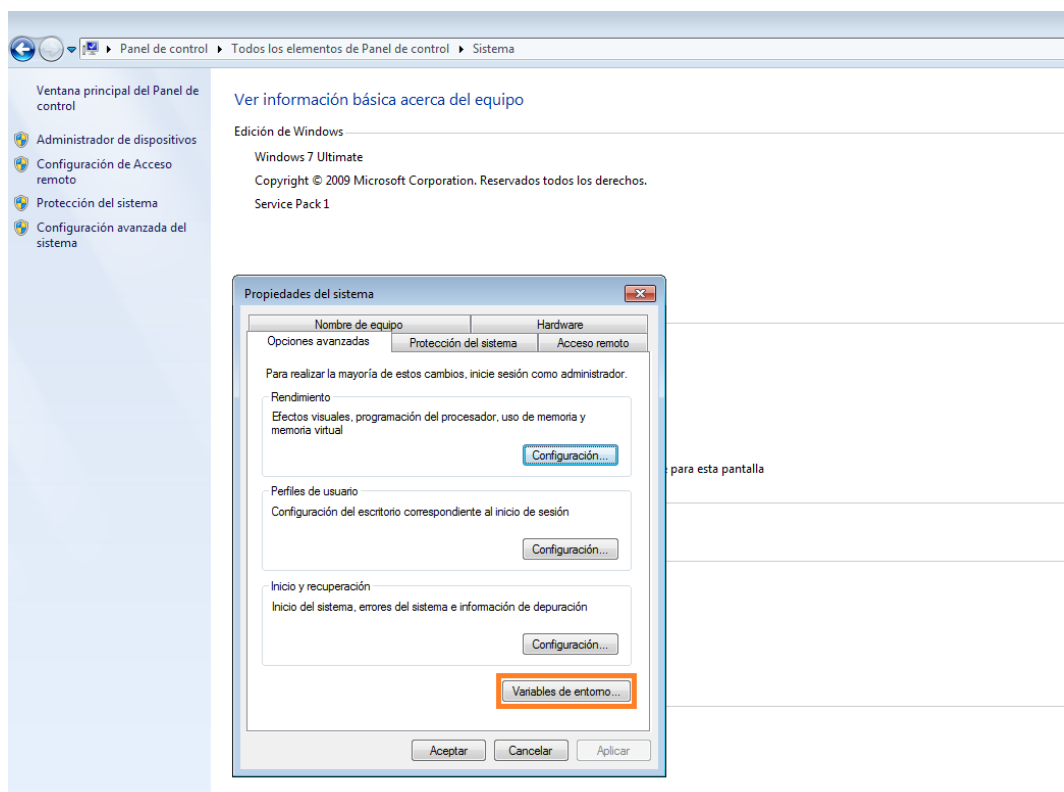


Fig. 2. Propiedades del sistema

Una vez cliquemos en la pestaña *Variables de entorno* se abrirá otra ventana donde se encuentran las *Variables de usuario* y las *Variables del sistema*. En esta última buscamos una variable llamada *Path* y la editamos para añadir lo siguiente: `;C:\Python27;C:\Python27\Scripts`. (Véase Figura 3)

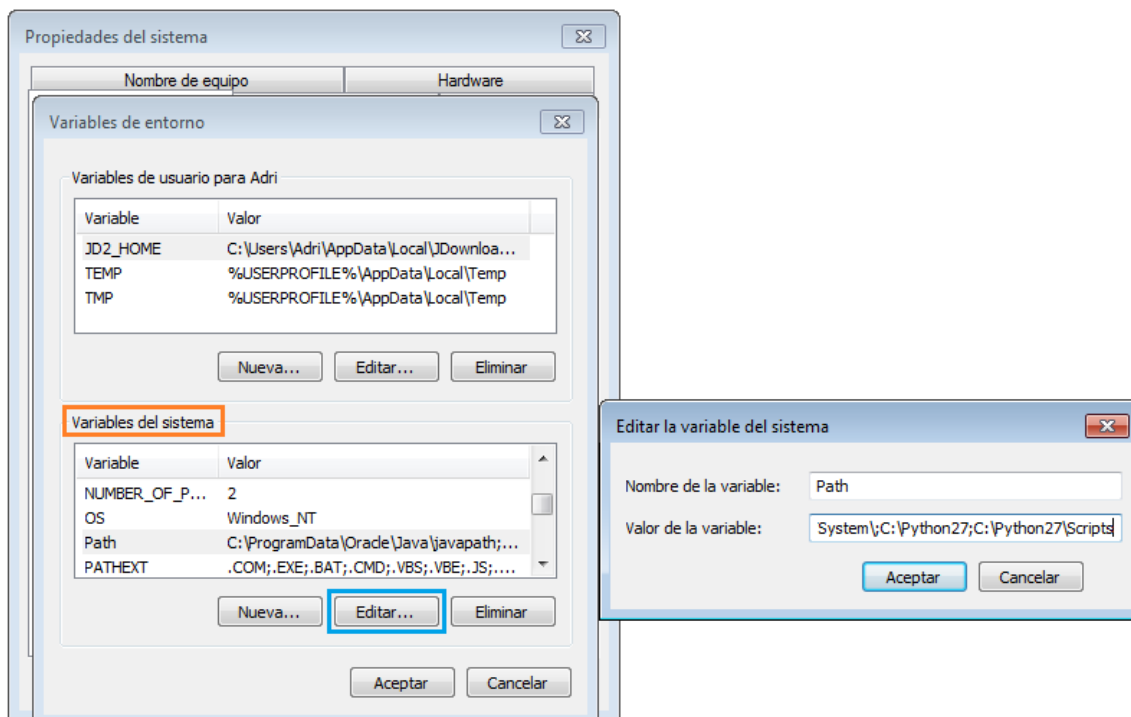


Fig. 3. Editar la variable Path

Finalizada la instalación de Python nos dirigimos a la carpeta llamada Instalador y hacemos doble clic en el ejecutable que tiene ese mismo nombre. Acto seguido, se abrirá una ventana como la de la Figura 4 y seguimos los pasos ya indicados por el ejecutable con el fin de instalar los paquetes necesarios para la ejecución del Source Seeker. Es necesario saber el sistema operativo del que se dispone y la versión de Python para instalar los paquetes de Windows 32 o Windows 64. En este caso instalaremos Windows 32 bits.

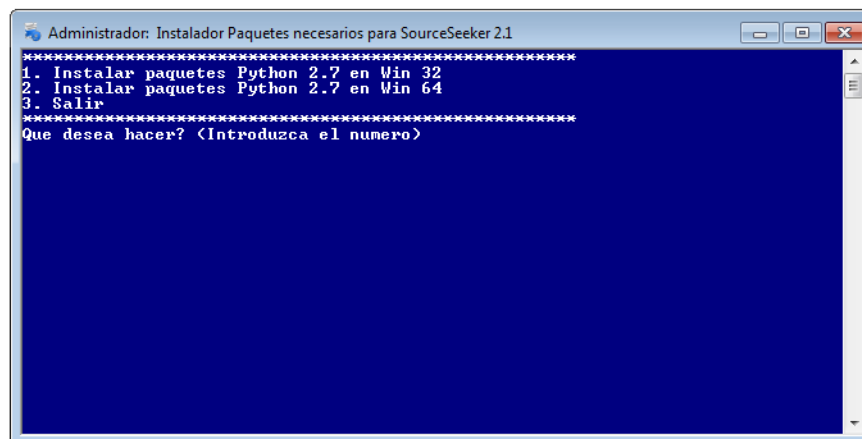


Fig. 4. Interfaz Instalador de paquetes

Finalmente, hacemos doble clic en el ejecutable *Source Seeker* y el programa se abrirá. (Véase Figura 5).

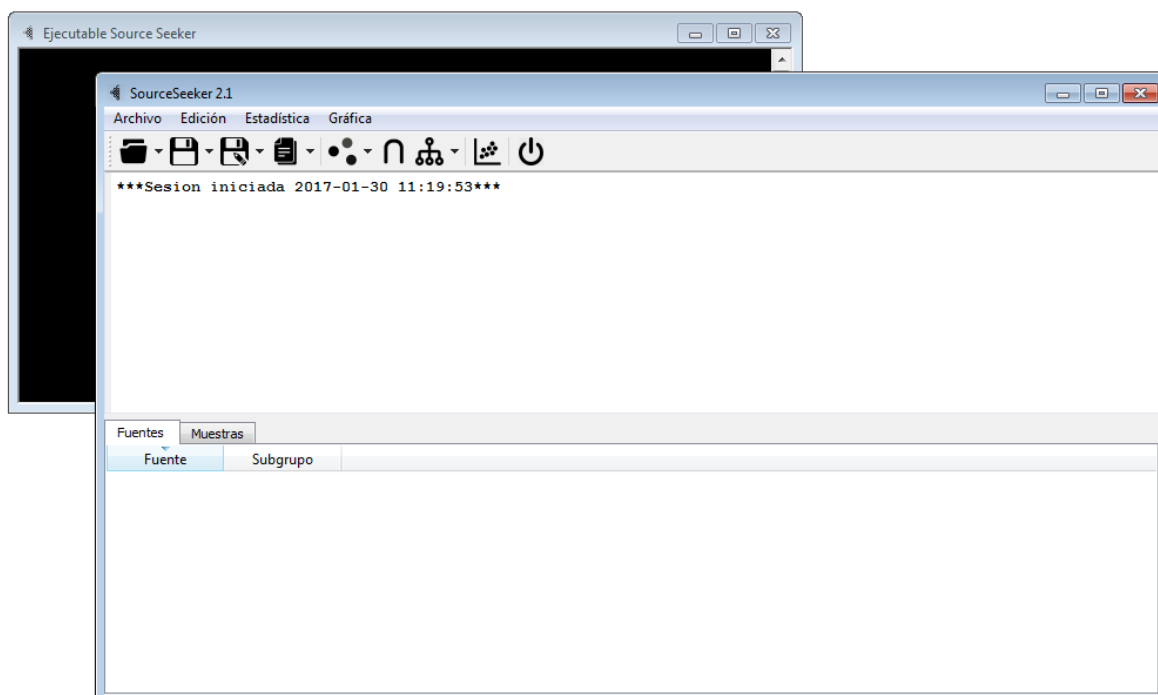


Fig. 5. Ejecutable del programa e interfaz principal