

# Selección de AGN utilizando criterios de color en el infrarrojo

Cedric Prieels

10 de febrero 2017

## Resumen

Esta práctica tiene por objetivo principal el estudio de un survey hecho por el cartografiador infrarrojo SWIRE usando el software TopCat, para seleccionar candidatos de AGN. Esta selección se hace utilizando unos criterios de color en el infrarrojo, porque sabemos que se pueden separar los AGN de la mayoría de las otras fuentes astronómicas simplemente representando este índice de color en un diagrama color-color. También queremos poder identificar en este ejercicio los contaminantes después de la selección, estimar la eficiencia y poder comprobar nuestra selección con la herramienta NED (*NASA Extragalactic Database*) de Caltech.

## 1. Introducción teórica

Como lo estudiamos en clase, los AGN (Active Galactic Nuclei) son objetos de gran interés en astronomía, que tienen una luminosidad y una densidad muy elevadas. Pensamos que estos AGN se forman por acrecimiento de materia por un agujero negro supermasivo situado en el centro de una galaxia, porque el disco de acreción que se forma de esta manera se puede calentar por rozamiento hasta llegar a temperaturas muy altas, del orden de millones de Kelvin. Este disco emite por lo tanto fotones en el ultravioleta y rayos X muy energéticos, que calientan el entorno del AGN, constituido de gas y polvo en forma de toro, y que empieza por lo tanto a emitir el mismo fotones en el infrarrojo que podemos detectar y analizar.

Sabemos hoy en día que existen diferentes tipos de AGN (galaxias tipo Seyfert, agujeros negros supermasivos, blazeres,...) pero las fuentes que nos van a interesar en este informe son los cuásares. Estos cuásares son objetos particulares en el sentido que son muy luminosos, y que se pueden por lo tanto observar desde muy lejos. Estos objetos son también interesantes a estudiar porque tienen una particularidad : sus índices de color en el infrarrojo cambian muy poco con el redshift  $z$ , porque la relación entre longitud de onda y el flujo (la *SED*, *Spectral Energy Distribution*) es una ley de potencia si consideramos fotones en el infrarrojo. Se puede por lo tanto representar la evolución del índice de color de cuásares en función del redshift, y veremos que la traza obtenida es particular y nos da un corte interesante que podemos aplicar para seleccionar cuásares dentro de una lista conteniendo diferentes objetos astronómicos.

Para estudiar y seleccionar los objetos que nos interesan, vamos a usar una pequeña muestra que proviene de un survey hecho por SWIRE (*Spitzer Wide-area InfraRed Extragalactic*), un detector de gran area (unos 50 grados cuadrados) capaz de considerar unas  $10^6$  fuentes en 4 bandas del infrarrojo solamente observables desde el espacio (a 3.6, 4.8, 5.8 y 8  $\mu m$ ). Este detector nos resulta muy útil para buscar AGN, porque puede detectar directamente picos de emisión de luz debidos al toro de gas y polvo que entorna las AGN y que emite en el infrarrojo. Este telescopio es capaz de medir solamente el flujo de los objetos en Jansky, y tendremos por lo tanto que calcular nosotros la magnitud en el sistema AB y el índice de color a partir de este flujo usando la relación siguiente

$$m_{AB} = 16,4 - 2,5\log(F) \quad (1)$$

El índice de color se calcula fácilmente después porque se define como la diferencia entre las magnitudes observadas en diferentes bandas, y nos permitirá seleccionar candidatos a ser cuásares a partir de la base de datos que tenemos.

La práctica consiste en seleccionar y imponer una serie de cortes que nos van a permitir seleccionar en esta base de datos objetos de interés. Una vez hecho, tendremos que estimar los contaminantes de nuestra selección (con diferentes métodos, como se explicará en la metodología) y tendremos que calcular la eficiencia y la completitud de la selección aplicada. La eficiencia de la selección se calcula considerando el número de candidatos seleccionados que resultan ser AGN de verdad, mientras que la completitud se define como la fracción de AGN verdaderos que seleccionamos. Por supuesto, queremos obtener una eficiencia y una completitud lo más alto posible para tener una selección más precisa pero nunca vamos a poder tener estos dos números muy altas en el mismo tiempo (por ejemplo, aumentar la eficiencia de la selección implica que la completitud irá bajando).

## 2. Metodología

Tenemos a nuestra disposición un survey hecho por SWIRE y que contiene un número de objetos muy grande. Lo primero que hacemos consiste en elegir una pequeña área del cielo para seleccionar una parte solamente de estos objetos del survey. Después, con esta submuestra, vamos a tener que convertir el flujo medido por el telescopio a magnitudes en el sistema AB usando la función *janskyToAb*( ) de TopCat, para después poder calcular el índice de color de cada objeto. Este cambio se realiza para las 4 bandas en el infrarrojo que nos interesan. Hay que estar cuidado que si el flujo emitido por una fuente está por debajo del umbral de detección del detector, SWIRE considera este objeto como teniendo un flujo nulo, y por lo tanto una magnitud en el sistema AB infinita.

Ahora que tenemos acceso a las magnitudes en el sistema AB en nuevas columnas, podemos empezar a calcular el índice de color de cada objeto, que se define como la diferencia entre las magnitudes de una fuente en diferentes bandas. Por ejemplo, el índice de color que llamaremos  $c_{12}$  se calcula como siendo la diferencia entre la magnitud en la banda de  $3.6\mu m$  y la magnitud en la banda de  $4.5\mu m$ . Podemos pintar todos los índices de color obtenidos de esta manera en unos diagramas color-color.

Una vez hecho, podemos empezar la selección de objetos de interés. Lo primero que hacemos consiste en pintar la traza de diferentes cuásares a partir de un otro survey hecho por SWIRE que contiene solamente cuásares, para poder determinar en qué parte del diagrama color-color se sitúan en general los cuásares, y por lo tanto una serie de cortes de selección. El objetivo consiste en encontrar uno(s) corte(s) que nos van a permitir rechazar todos los objetos que tienen colores lejanos de las trazas de los cuásares que acabamos de obtener.

Esta selección nos permite reducir el número de candidatos AGN a un número razonable, pero todavía tenemos que encontrar un método que nos permite estimar el número de contaminantes, es decir estimar el número de objetos del survey que vamos a considerar como cuásares pero que no lo son (pero que tienen un índice de color cerca al índice típico de cuásares). Dos métodos sencillos existen para estimar estos contaminantes :

- El **primer método** consiste en considerar otro(s) diagrama(s) color-color. Un objeto que no sea un cuásar puede quedar cerca de la zona de selección del diagrama color-color que hemos considerado hasta ahora por casualidad pero no debería volver a estar cerca de esta zona si consideramos otro diagrama color-color. Por lo tanto, podríamos primero hacer un *subset* para seleccionar solamente del survey los objetos que están dentro de nuestra zona de selección de la figura 5, y después, pintar el mismo plot con este *subset* pero considerando otros colores.

- El **segundo método** consiste en considerar solamente AGN de tipo 1, que son fuentes puntuales. Consideramos por lo tanto la columna *mst* del survey, que nos indique si la fuente observada es puntal o extendida y podemos de esta manera descartar de nuestra selección todos los objetos extendidos (galaxias) con un simple *subset*.

Por fin, solo hace falta calcular la eficiencia y la completitud de nuestra selección. Esta etapa es necesaria porque nuestra selección no va a ser perfecta (vamos a perder algunos AGN y seleccionar objetos del survey que no son AGN verdaderos). Tenemos también al final que verificar que los objetos que hemos seleccionados son cuásares de verdad, gracias a la herramienta NED.

### 3. Resultados obtenidos

Para realizar esta práctica, se selecciona primero dentro del software TopCat una submuestra del survey total considerando un valor de ascensión recta de 16 grados y 10 minutos, un valor de declinación de 55 grados y 20 minutos y un campo de 0.56 grados.

#### 3.1. Cálculo de los índices de color

Una vez hecho, abrimos esta muestra de fuentes y calculamos como explicado en la metodología los valores de magnitudes en el sistema AB a partir de los flujos medidos en cada banda del infrarrojo, y guardamos estos valores en nuevas columnas. Después, calculamos todos los índices de color posibles haciendo la diferencia entre las magnitudes que acabamos de calcular. Estos últimos resultados también se guardan en nuevas columnas. Con estos índices de color, podemos empezar a pintar los diagramas color-color que queremos.

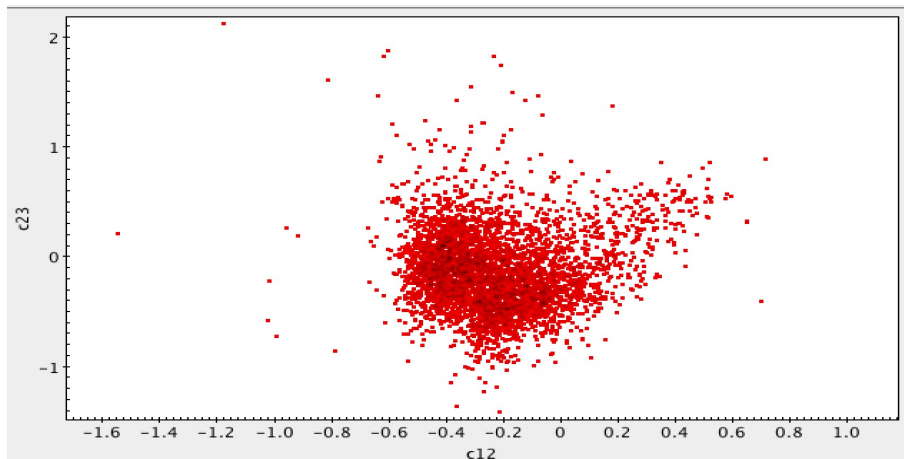


Figura 1: Ejemplo de diagrama de color representado el índice de color  $[3.6]-[4.5]$  vs el índice  $[4.5]-[5.8]$

#### 3.2. Estudio del color de diferentes objetos

Ahora, podemos pintar la traza de los cuásares (evolución del índice de color de los cuásares en función del redshift  $z$ ), usando TopCat y la tabla de datos a nuestra disposición en un fichero txt. El plot resultando se enseña en la figura (2), en el cuál vemos claramente que el índice de color depende del redshift (pero esta dependencia no es muy fuerte, como lo vimos en la introducción teórica).

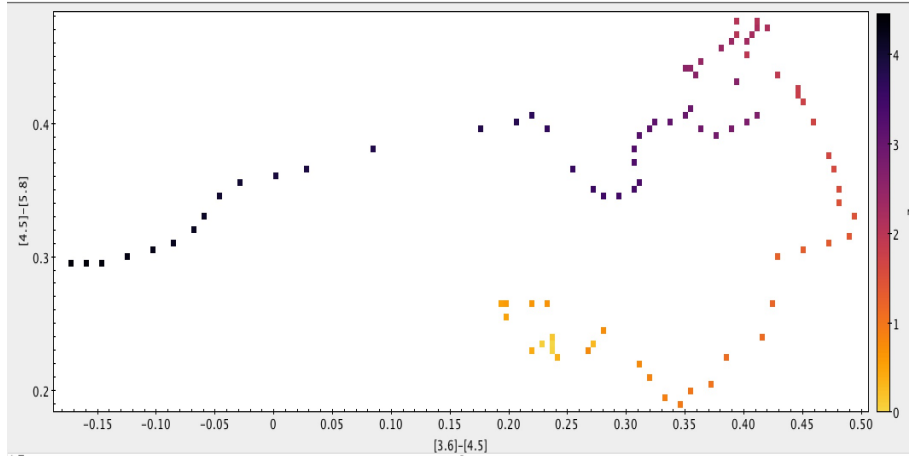


Figura 2: Traza de color del cuásar en función del redshift

Podemos también representar en el mismo plot la traza que acabamos de obtener para el cuásar que estudiamos y el índice de color de todos los objetos del survey de SWIRE. Este plot se representa en la figura (3), donde se observa directamente que vamos a poder seleccionar objetos candidatos a ser cuásares cortando en los valores de los índices de color.

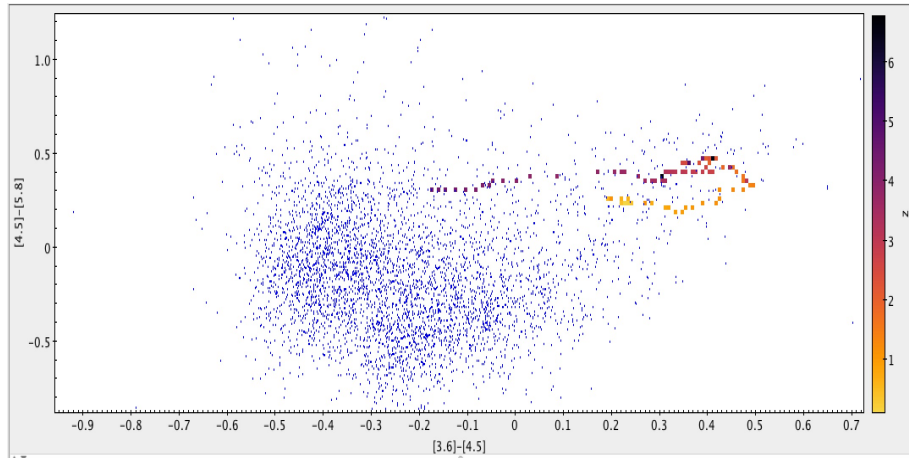


Figura 3: Índice de color de todos los objetos considerados del survey y traza de un cuásar

### 3.3. Correlación con catálogos de AGN conocidos

Ya que tenemos los plots representando dos índices de color para todos los objetos del survey y para un cuásar, podemos estudiar la variación de esta traza que hemos obtenido considerando diferentes cuásares. Este paso es por supuesto muy importante en la determinación del corte que vamos a aplicar para seleccionar candidatos a ser cuásares, porque necesitamos tener una idea de la dispersión de los colores de los cuásares. Para esto, abrimos en TopCat un survey de SWIRE constituido solamente de cuásares y añadimos los puntos así obtenidos en el plot anterior. Estos resultados se representan en la figura (4). Con esta figura, vemos ahora claramente que se pueden seleccionar los cuásares directamente poniendo unos cortes en los valores de los colores, porque los cuásares tienden a tener índices más altos y una dispersión bastante pequeña. En este figura, vemos también que nos va a resultar más fácil seleccionar los cuásares a bajo redshift que los otros, porque los cuásares a alto redshift parecen estar más mezclados con los puntos del survey de SWIRE.

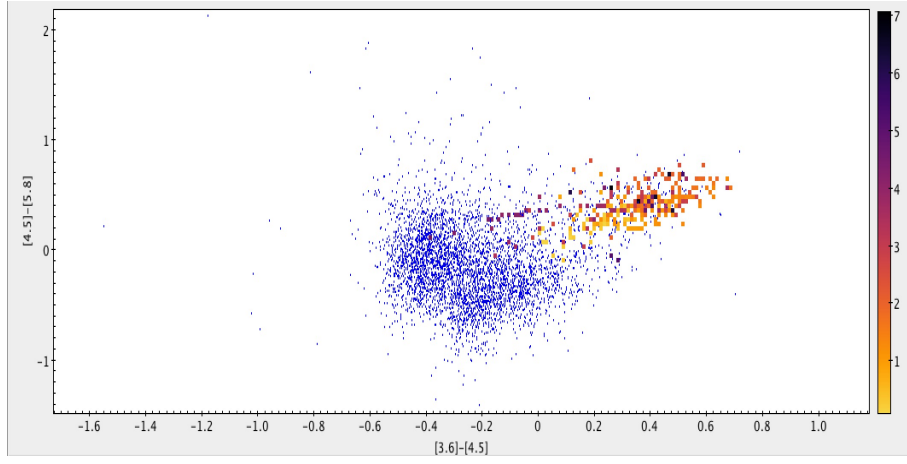


Figura 4: Estudio de la dispersión de los colores de diferentes cuásares en función del redshift

### 3.4. Selección de candidatos AGN

Podemos empezar la selección de candidatos AGN ahora. Como discutido antes, sería interesante separar el problema en dos partes distintas : primero, vamos a considerar todos los objetos y después, vamos a considerar solamente los objetos que tienen un redshift más débil (menor que 3.5).

#### Todos los cuásares

Ahora empezamos la selección considerando todos los cuásares, no solamente los que tienen un redshift mayor que 3.5. En este caso, la selección hecha es la siguiente :

$$\begin{cases} -0,1 < y < 0,8 \\ y < 1,8x + 0,6 \\ y > 1,8x - 0,7 \end{cases}$$

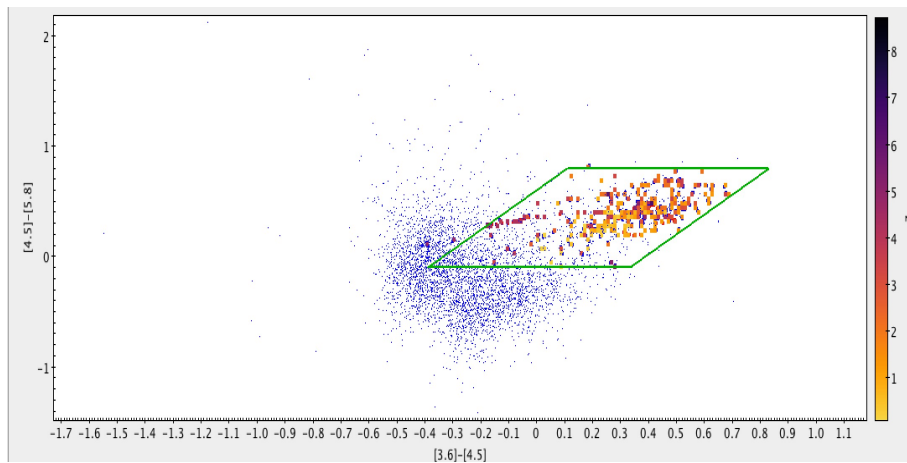


Figura 5: Selección considerando todos los cuásares

## Cuásares de redshift débil

Estos cuásares son más fáciles a caracterizar porque tienen en general colores muy diferentes de los colores medidos en el survey de SWIRE. En este caso, podemos usar el polígono definido por los cortes siguientes :

$$\begin{cases} -0,1 < y < 0,8 \\ y < 5x + 0,3 \\ y > 5x - 2,9 \end{cases}$$

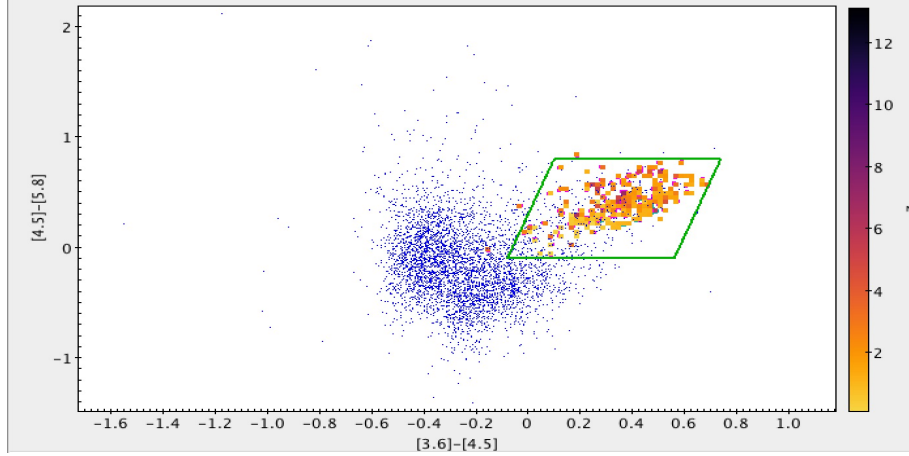


Figura 6: Selección considerando los cuásares de redshift menor que 4

### 3.5. Estimación de contaminantes

Como se explicó antes, con las dos series de corte que estamos aplicando, no vamos a seleccionar solamente AGN, porque solo estamos seleccionando candidatos a ser AGN. Tenemos por lo tanto que estimar el número de contaminantes de nuestra selección, es decir estimar el número de objetos del survey que vamos a considerar como AGN pero que no lo son (pero que tienen un índice de color cerca al índice típico de AGN). En este caso y en el resto del informe, vamos a considerar los cuásares que tienen cualquier valor de redshift (no solamente los más próximos).

Aplicando el segundo método explicado en la metodología (sabemos que los AGN tienen que ser puntuales y tenemos en la tabla de datos una columna llamada *mst* y que nos da esta información para cada objeto) y seleccionando por lo tanto solamente los objetos que tienen un *mst* igual a -1, reducimos nuestra selección hasta llegar a 59 candidatos (menos del 1% de los 27000 objetos que tenemos en la parte considerada del survey).

### 3.6. Eficiencia y completitud de la selección

Ahora que hemos estimado los contaminantes, podemos calcular la eficiencia y completitud de nuestra selección. La selección que hemos encontrado selecciona 59 candidatos de los 27000 que tenemos al principio en el survey, y selecciona 233 de los 278 AGN del catálogo de cuásares de SWIRE. Por lo tanto, podemos calcular la completitud  $c$  y la eficiencia  $\eta$  (si consideramos que uno de cada 1000 objetos del survey es un cuásar, y que por lo tanto esperamos tener más o menos 27 cuásares en la parte considerada del survey) :

$$\begin{cases} c = \frac{233}{278} = 0,838 \\ \eta = \frac{27_{\text{esperados}} \cdot 0,838}{59_{\text{encontrados}}} = 0,384 \end{cases}$$

### 3.7. Verificación de candidatos

Ahora que tenemos nuestra selección hecha, solo hace falta comprobar que los objetos que hemos seleccionado son cuásares de verdad, utilizando la herramienta NED de Caltech, que nos devuelve el tipo de un objeto a partir de su posición exacta en el cielo. Exportamos por lo tanto primero las posiciones de nuestros candidatos en Topcat, y los importamos en la base de datos NED. Los primeros 26 resultados encontrados dentro de un rayo de 2 segundos de arco se enseñan en la figura siguiente.

SOURCE LIST											
Objects are listed as the input order											
Row No.	Input Position / Object Name (E/N Hover for Messages)	Sep (")	Object Name (* Hover for Essential Note)	RA J2000.0	Dec J2000.0	Object Type	Redshift	z	Qual	Row No.	
1	# RAJ2000 DJ2000									1	
2	242.86613 54.942	0.3	ELAIS06(R) J161127.9+545631	16h11m27.91s	+54d56m31.3s	G				2	
3	241.96922 54.95043	0.2	ELAIS05(R) J160752.6+545701	16h07m52.61s	+54d57m01.7s	G				3	
4	242.26865 54.97276	0.4	ELAIS05(R) J160904.5+545822	16h09m04.51s	+54d58m22.2s	G				4	
5	242.13684 54.97321	0.0	SDSS J160832.84+545823.5	16h08m32.84s	+54d58m23.5s	G				5	
6	242.81605 54.98111	0.4	ELAIS06(R) J161115.8+545852	16h11m15.89s	+54d58m52.2s	VisS				6	
7	242.88563 55.02443	0.2	SDSS J161132.54+550127.7	16h11m32.54s	+55d01m27.7s	VisS				7	
8	241.9471 55.08128	0.1	SDSS J160747.31+550452.5	16h07m47.32s	+55d04m52.6s	G				8	
9	243.04733 55.11513	0.1	SSTSL2 J161211.35+550654.5	16h12m11.35s	+55d06m54.5s	IrS				9	
10	243.10243 55.1362	0.1	SDSS J161224.58+550810.2	16h12m24.58s	+55d08m10.2s	G				10	
11	243.23225 55.14684	0.1	SDSS J161255.97+550848.5	16h12m55.98s	+55d08m48.5s	G				11	
12	241.94794 55.15237	0.2	SSTSL2 J160747.51+550908.3	16h07m47.52s	+55d09m08.3s	IrS				12	
13	242.6237 55.16837	0.4	ELAIS02(R) J161029.7+551006	16h10m29.73s	+55d10m06.3s	G				13	
14	241.59862 55.17048	0.2	SSTSL2 J160623.69+551013.6	16h06m23.70s	+55d10m13.7s	IrS				14	
15	241.8217 55.1763	0.2	SDSS J160717.22+551034.5	16h07m17.23s	+55d10m34.5s	G				15	
16	242.21481 55.17676	1.2	ELAIS02(R) J160851.6+551036	16h08m51.69s	+55d10m36.5s	G				16	
17	242.40665 55.20341	0.3	SDSS J160937.62+551212.4	16h09m37.62s	+55d12m12.4s	VisS				17	
18	243.17981 55.21867	0.2	SSTSL2 J161243.17+551307.0	16h12m43.18s	+55d13m07.1s	IrS				18	
19	242.24998 55.27136	0.5	ELAIS02(R) J160900.0+551616	16h09m00.05s	+55d16m16.7s	G				19	
20	242.33536 55.2743	0.2	SDSS J160920.51+551627.4	16h09m20.51s	+55d16m27.4s	G				20	
21	242.07759 55.2725	0.4	ELAIS02(R) J160818.6+551620	16h08m18.67s	+55d16m20.9s	G				21	
22	243.34373 55.28237	0.4	SSTSL2 J161322.54+551656.5	16h13m22.55s	+55d16m56.5s	IrS				22	
23	243.08032 55.35301	0.1	SSTSL2 J161219.28+552110.9	16h12m19.28s	+55d21m10.9s	IrS				23	
24	241.5269 55.37764	0.3	SDSS J160606.48+552239.2	16h06m06.48s	+55d22m39.2s	G				24	
25	243.00934 55.37922	0.3	SDSS J161202.26+552244.9	16h12m02.27s	+55d22m44.9s	VisS				25	
26	243.28085 55.40229	0.5	SSTSL2 J161307.45+552408.3	16h13m07.46s	+55d24m08.3s	IrS				26	

Figura 7: Primeros 26 resultados encontrados en la base de datos NED

Con la selección determinada a lo largo del informe, vemos que desafortunadamente ninguno de los objetos seleccionados parece haber sido identificado por el detector, porque ningún objeto tiene valor de redshift asignado (solo un tipo de objeto, y no parece que ninguno sea un cuásar). Tendríamos que repetir la práctica considerando una otra parte distinta del cielo para ver si estos resultados se confirman o no.

## 4. Conclusión

Esta práctica nos permitió usar los datos de un survey hecho por el cartografiador infrarrojo SWIRE para seleccionar candidatos de AGN, imponiendo una serie de cortes sobre todos los objetos (seleccionamos solamente los objetos puntuales que tenían un índice de color cercano al índice de color típico de un cuásar).

Con la selección aplicada, hemos seleccionado 59 candidatos dentro de los 27000 objetos de la parte del survey elegida, con una completitud de 0,838 y una eficiencia de selección de 0,384. Hemos intentado comprobar con la herramienta NED si estos candidatos son cuásares de verdad o no, pero parece que ninguno de estos objetos ha sido identificado por el telescopio (por lo menos, ninguno tiene valor de redshift asociado).