

The logo for the XXV Congreso Estatal de Astronomía, consisting of the letters 'CEEA' in a stylized, white, sans-serif font. The 'C' is a simple circle, the 'E' has three horizontal bars, and the 'A' has a dot on its right vertical stroke.

CEEA

XXV CONGRESO ESTATAL
DE ASTRONOMÍA

ZARAGOZA
28 - 30 abril 2023

Aprendizaje automático en catálogos astronómicos

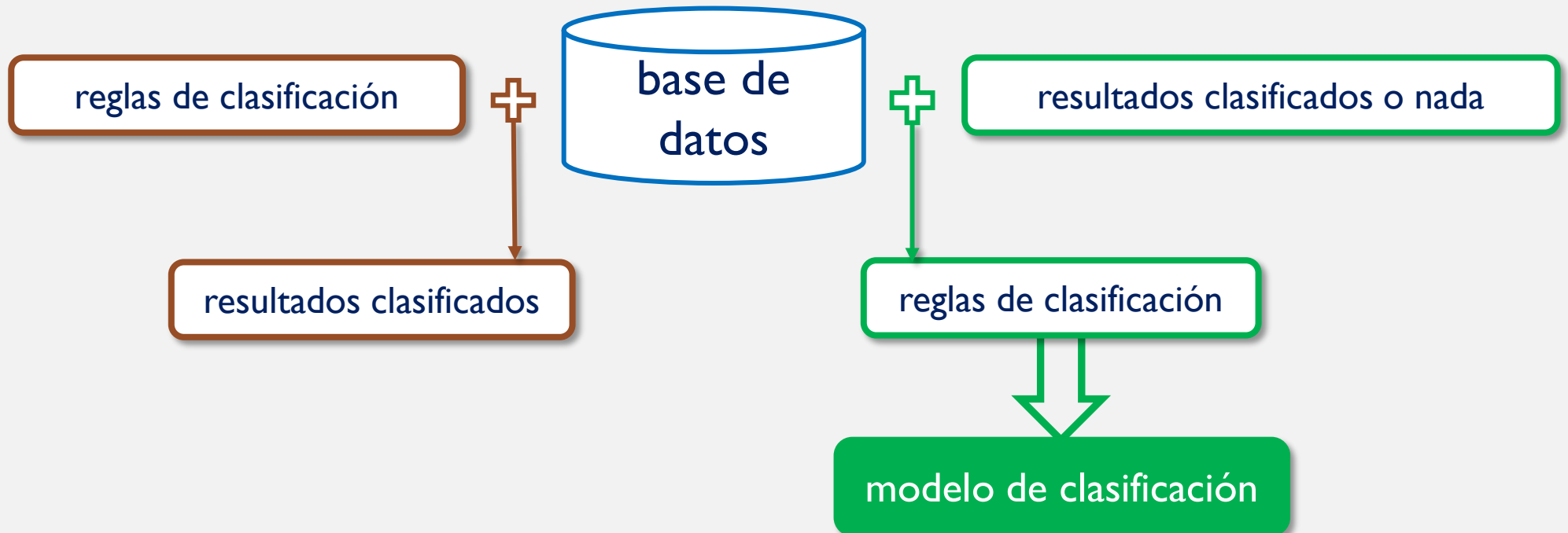
Esteban Fernández Mañanes



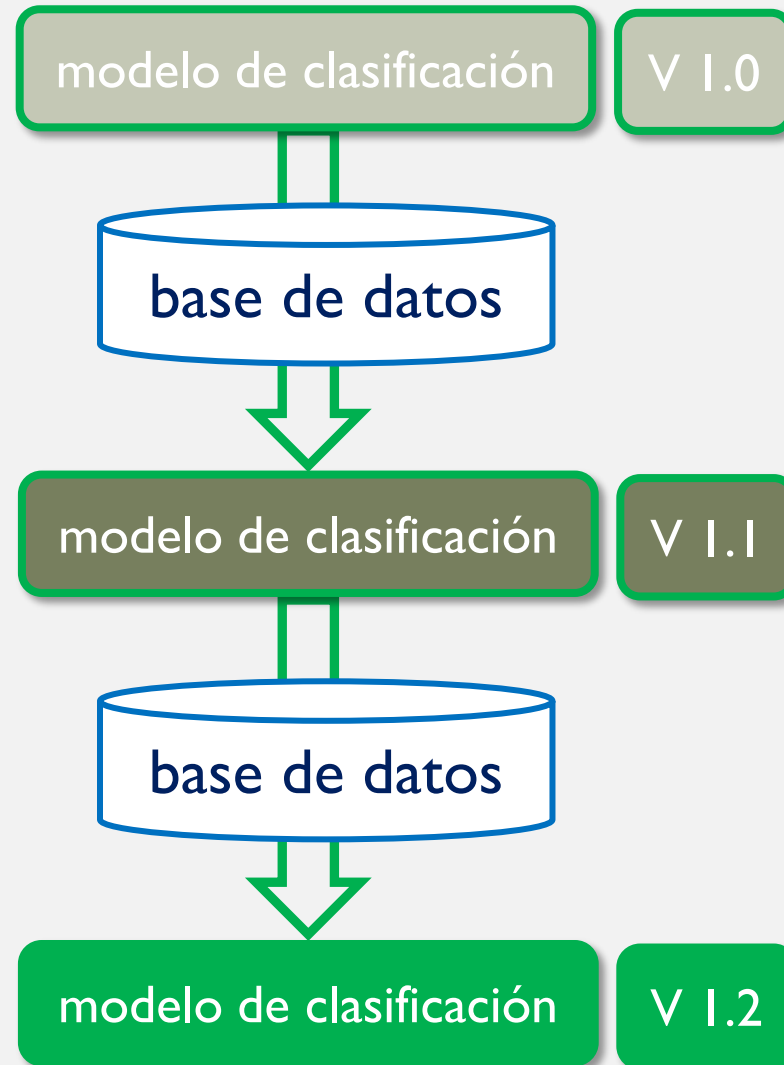
¿qué es el aprendizaje automático?

programación clásica

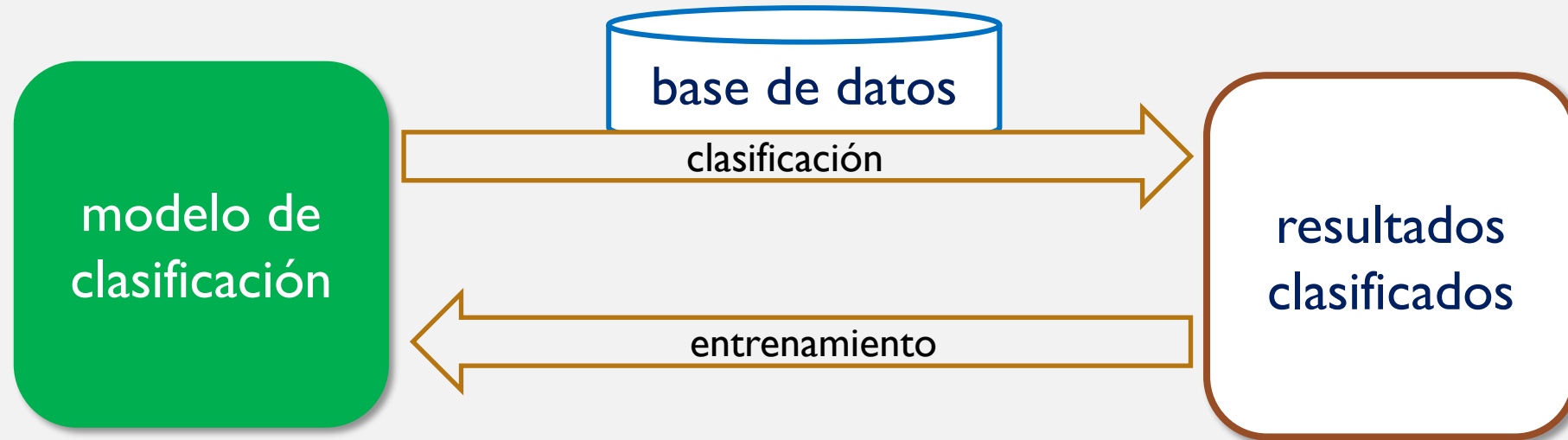
aprendizaje automático



entrenando el modelo



clasificación automática



valores en millones

DR1

DR2

EDR3

1 parámetros

1.142

1.693

48%

1.812

59%

2 parámetros

1.140

361

344

5 parámetros

2

1.331

585

6 parámetros

882

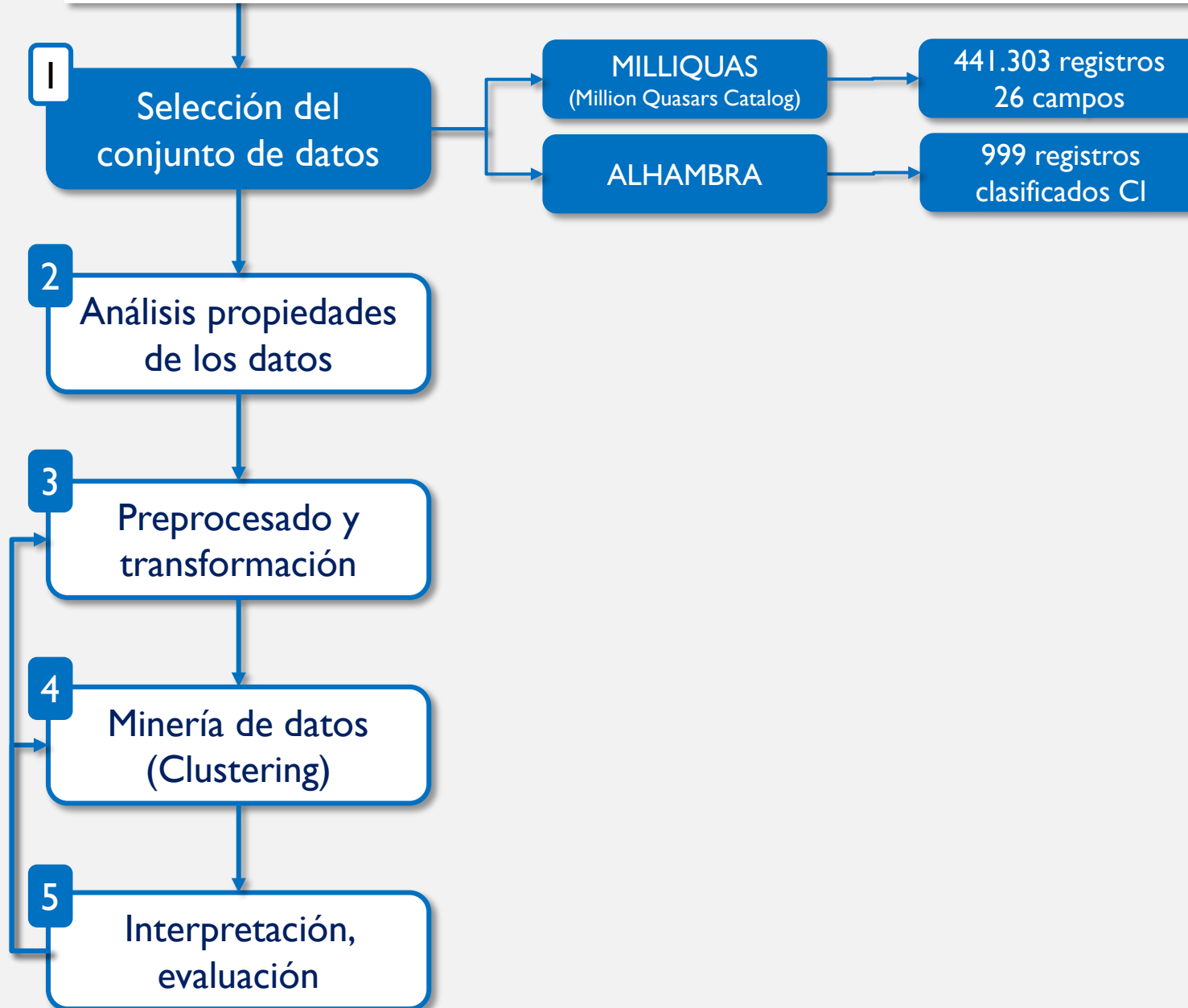
Clasificados
automáticos

0,003

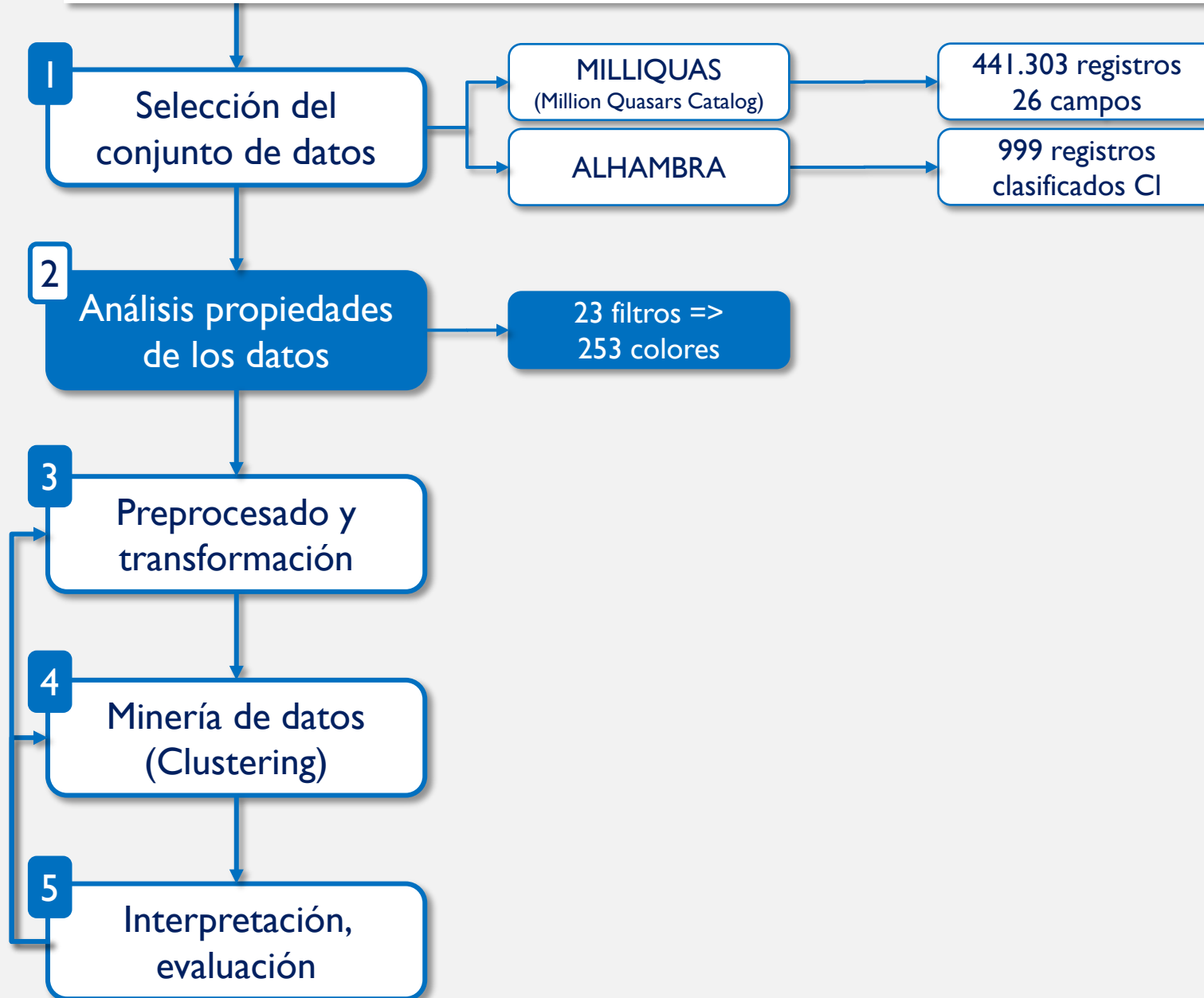
0,550

10,5

Fases de KDD_ Descubrimiento de conocimiento en las bases de datos



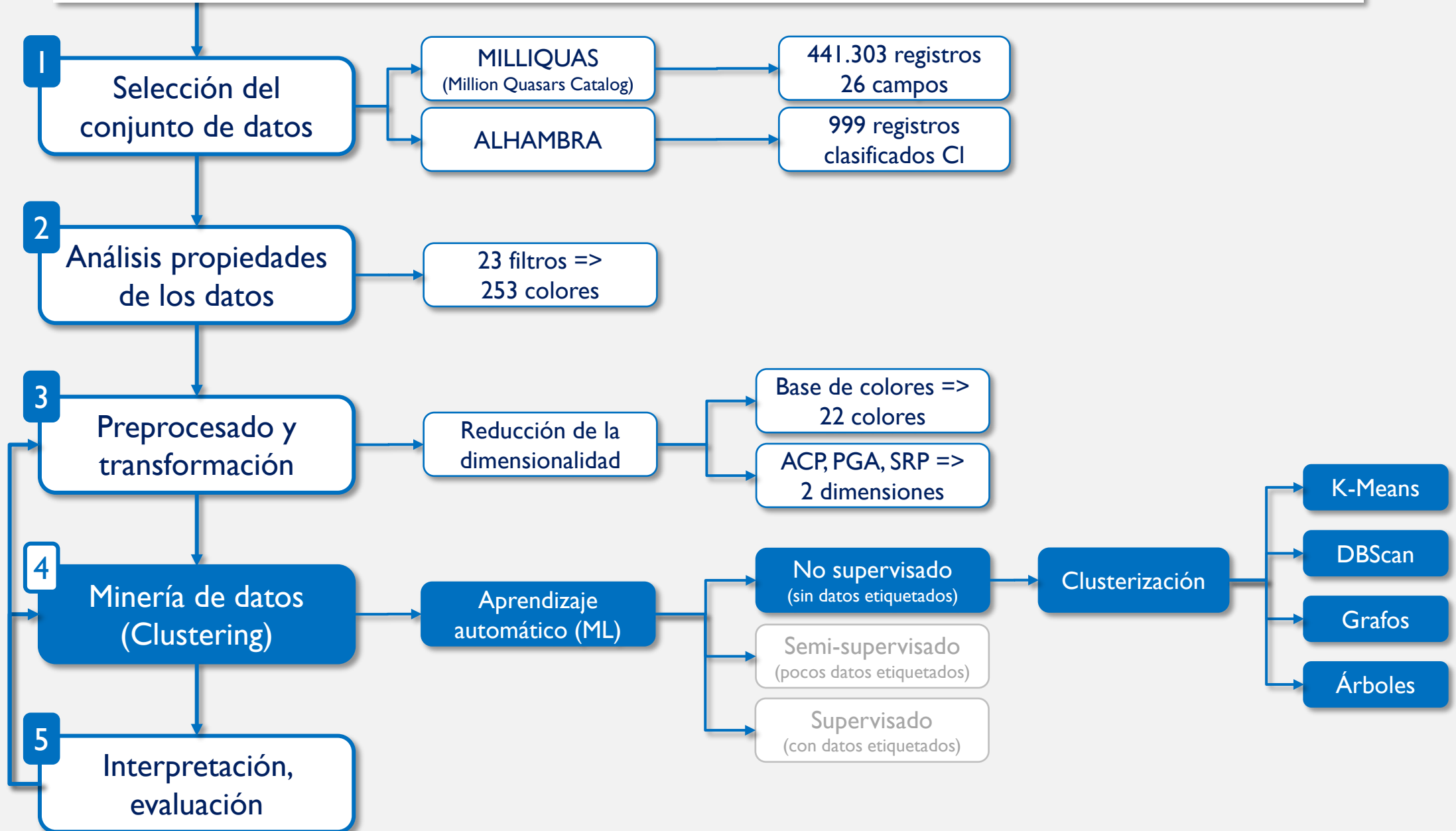
Fases del KDD_ Descubrimiento de conocimiento en las bases de datos



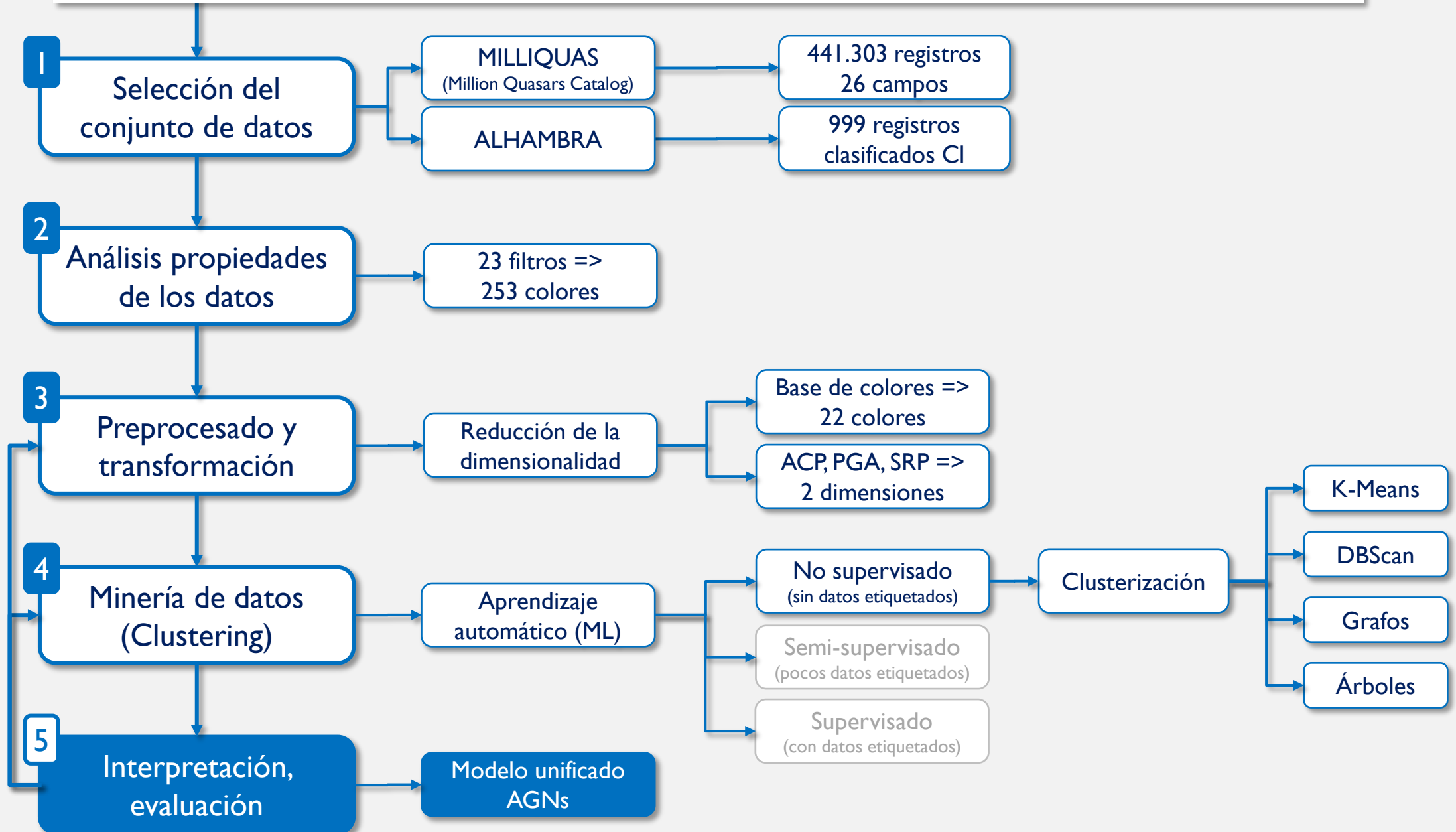
Fases del KDD_ Descubrimiento de conocimiento en las bases de datos



Fases del KDD_ Descubrimiento de conocimiento en las bases de datos



Fases KDD_ Descubrimiento de conocimiento en las bases de datos



Preprocesamiento de datos

Cambio de escala y normalización

Escala exponencial a lineal (euclídea).

- Estandarización
- Normalización

La maldición de la dimensionalidad

Problemas de la dimensionalidad

- Aumento de la variabilidad

Desaparecen las agrupaciones

Aumento de tiempo de CPU y memoria

La maldición de la dimensionalidad

	Registros	Dimensiones	Datos únicos	Información	Reducción de datos únicos
Datos iniciales	432.687	253	103.488.160	100 %	
Datos con base de colores	432.687	22	9.519.114	100 %	90 %
Datos reducidos <small>PCA, PGA o SRP</small>	432.687	2	865.374	84 %	91 %

0,85 % de los datos iniciales

84 % de la información original

2 dimensiones

Clasificar AGNs | Modelo unificado

Grupo 1

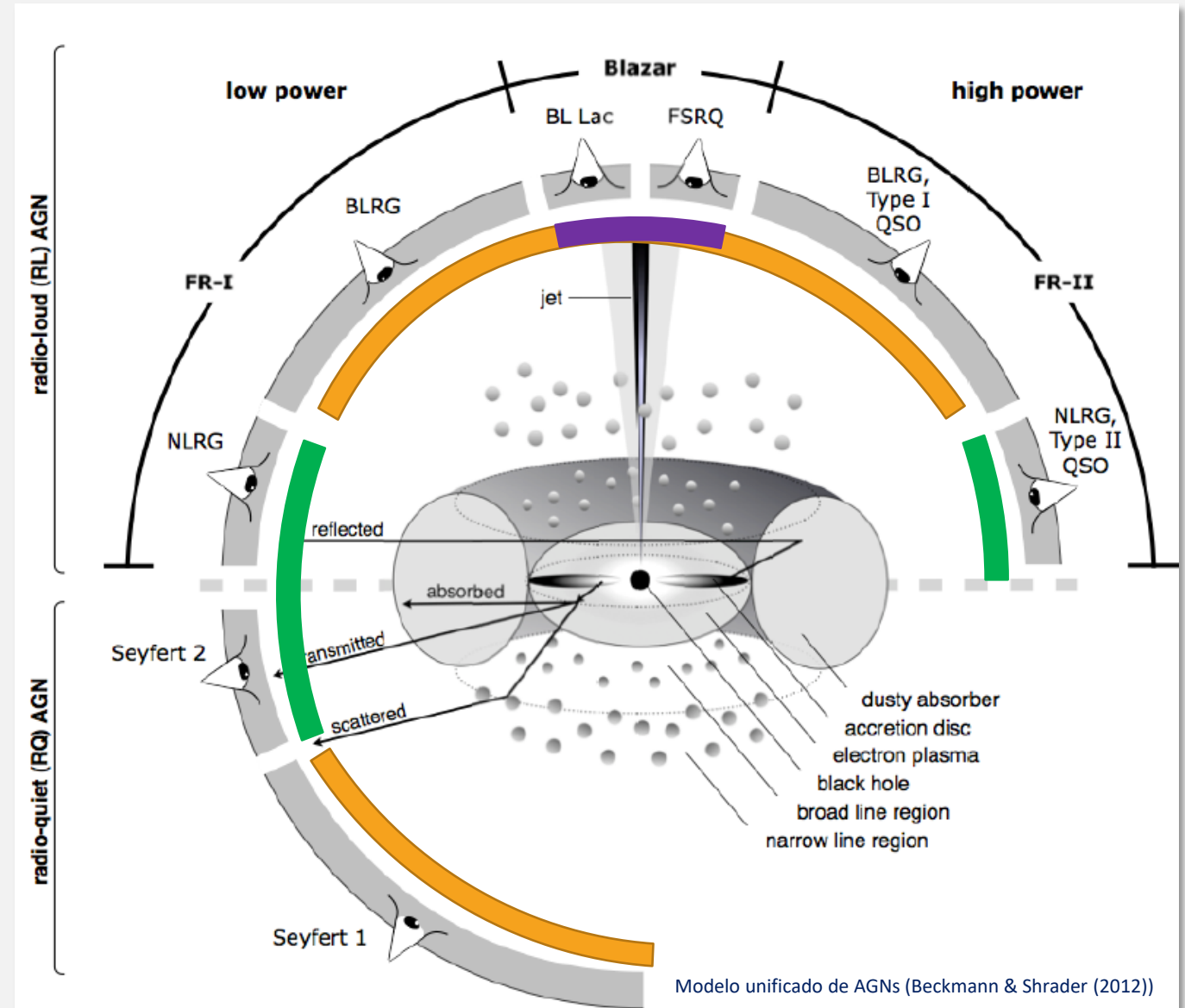
Líneas de emisión anchas y prominentes

Grupo 2

Líneas de emisión estrechas

IR, Rayos X más prominente que Grupo 1

Grupo 0



Clasificar AGNs | Modelo unificado

Modelo unificado

Orientación

Cuáasar Tipo I

CL 1

393

AGN Seyfert tipo I

CL 2

333

BL Lac

CL 3

0

Candidato cuáasar

CL 4

179

NELG Seyfert tipo 2

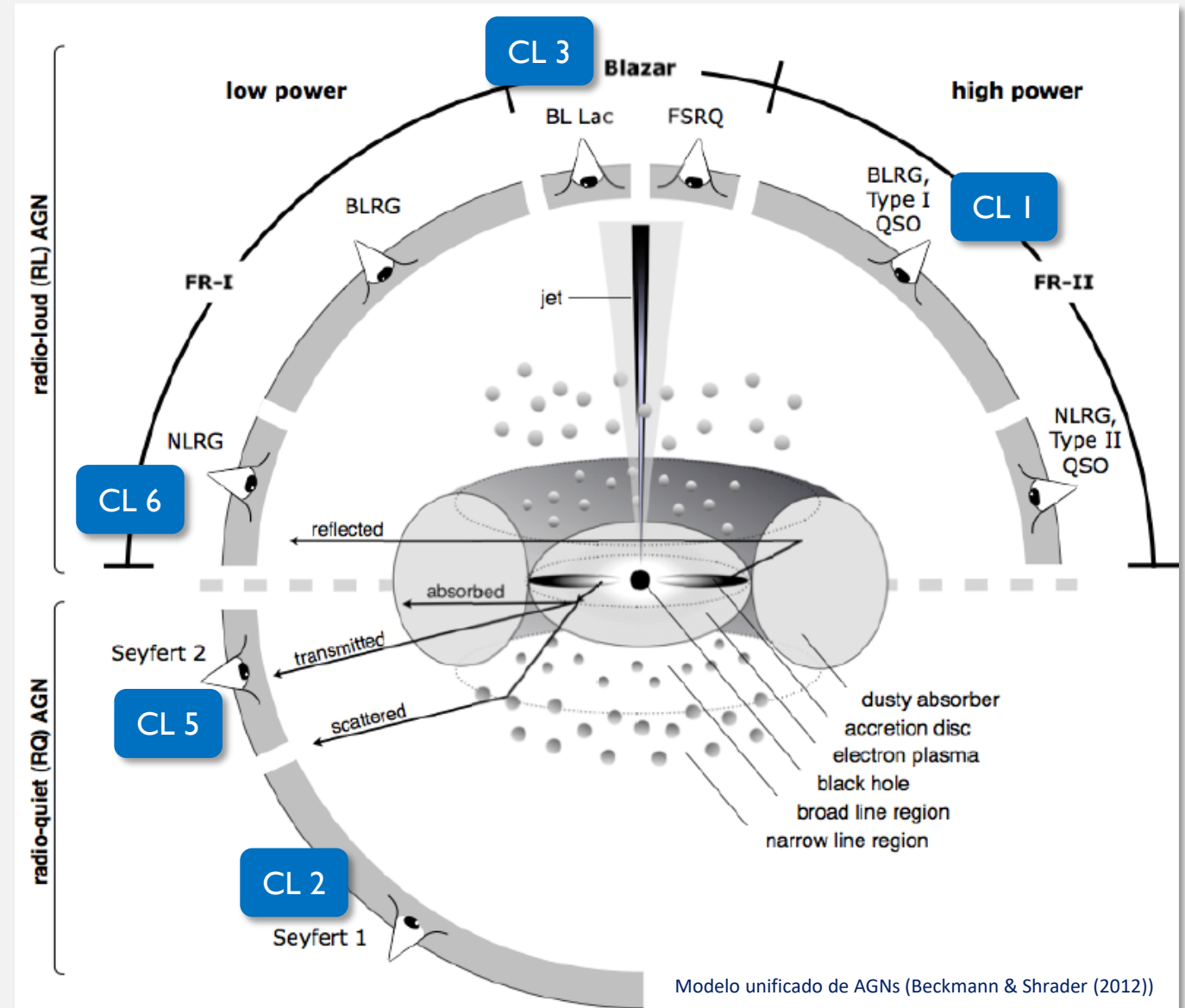
CL 5

81

NLAGN cuáasar

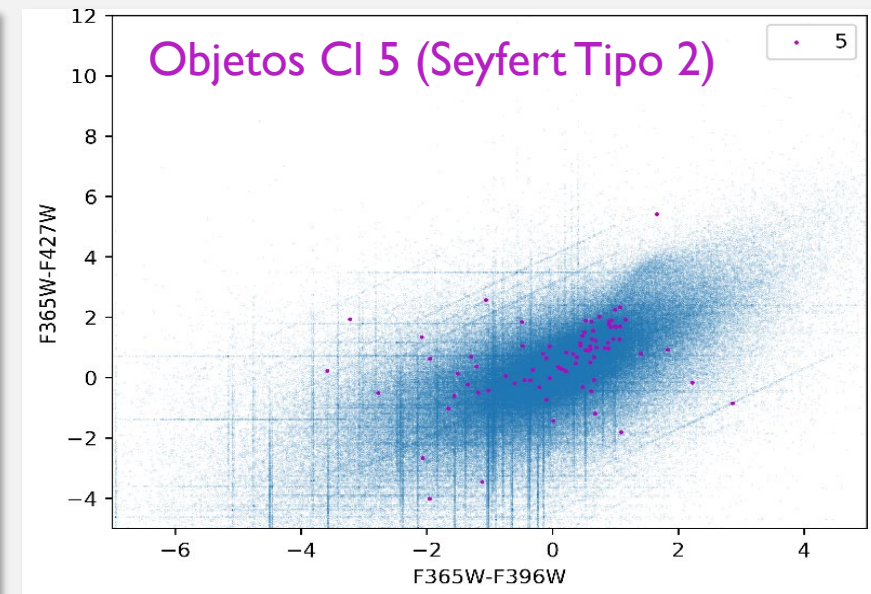
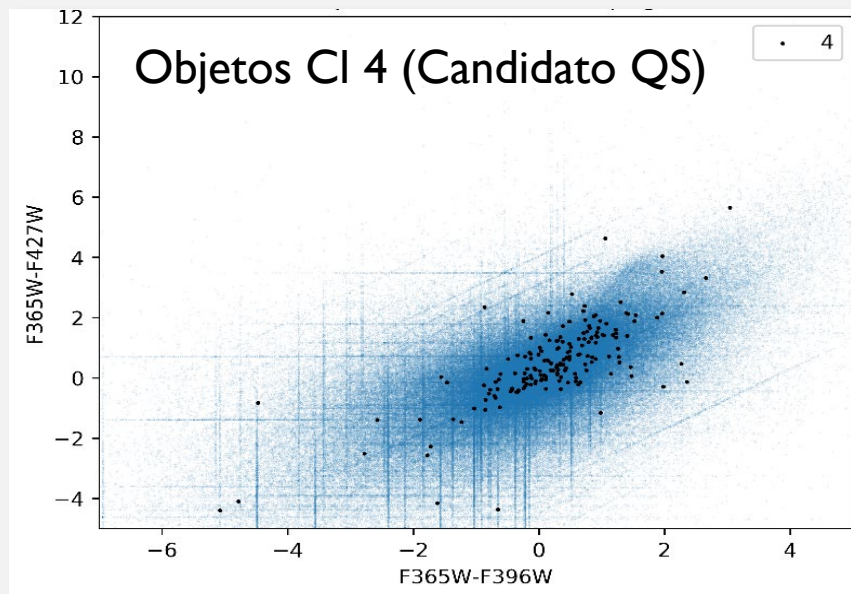
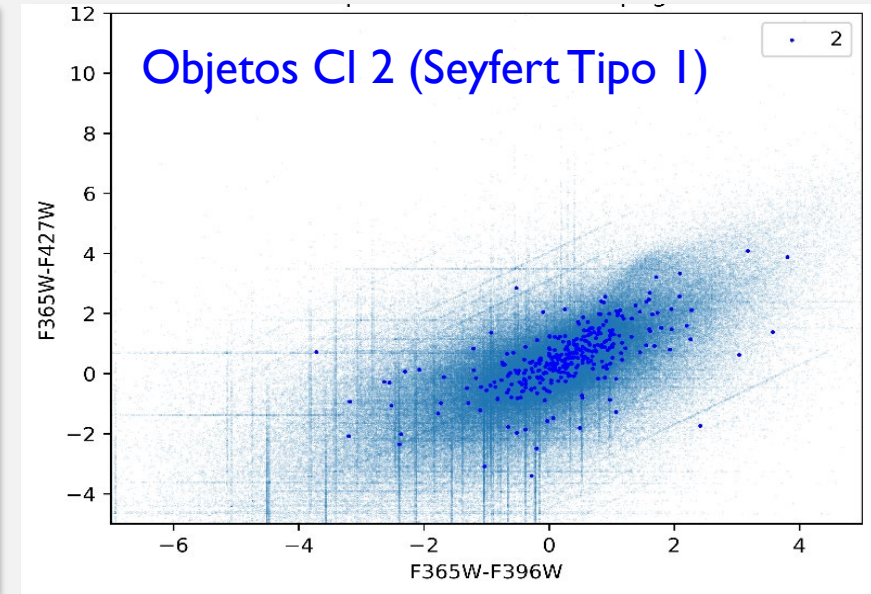
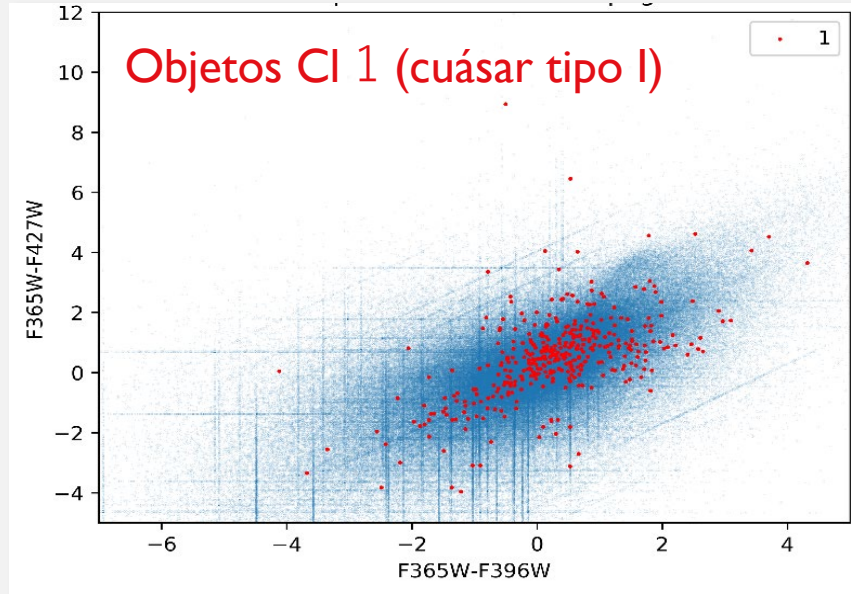
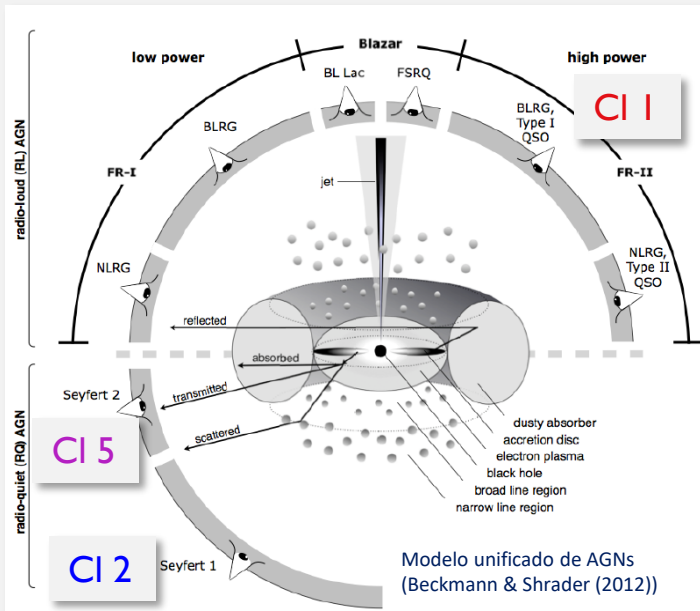
CL 6

13



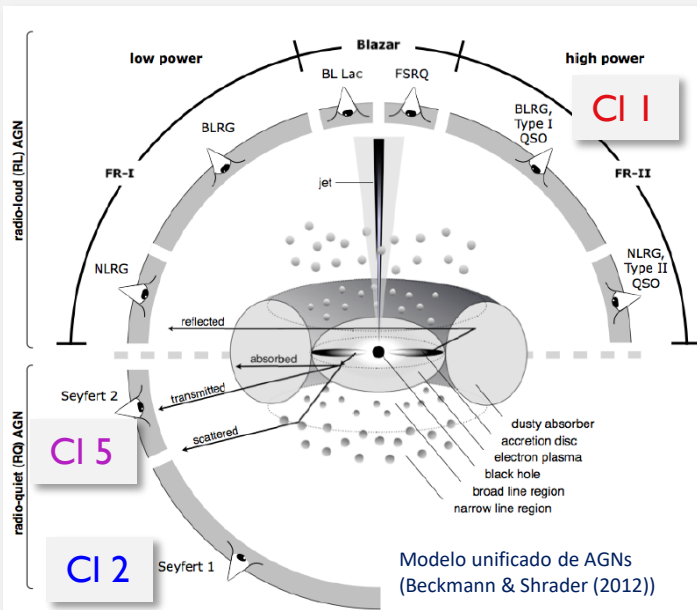
Dispersión color-color

- Gráfico de dispersión color-color
- Magnitudes lineales



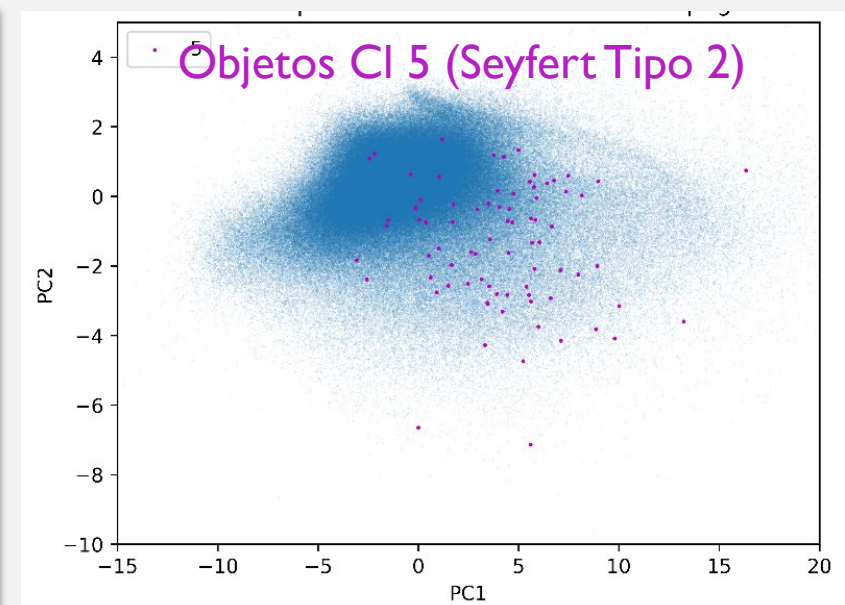
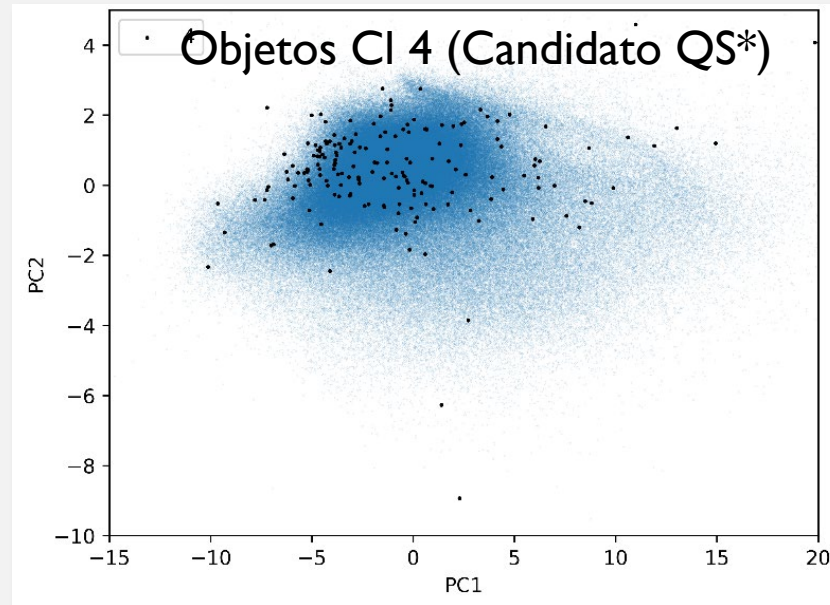
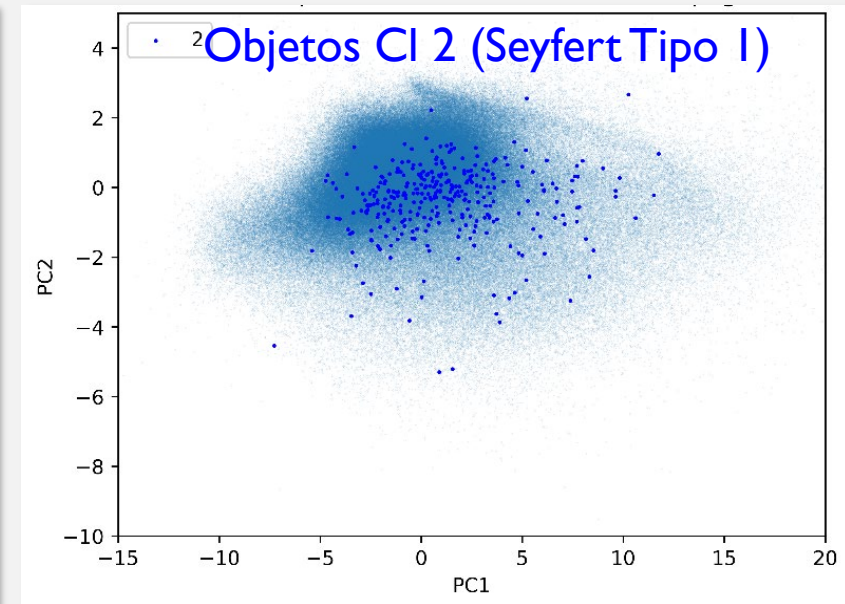
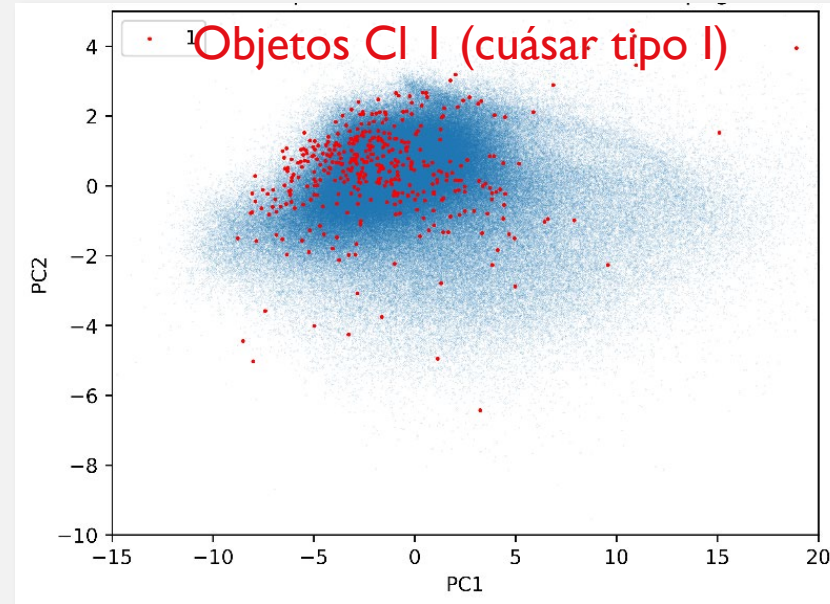
Espacio reducido por ACP

Se aprecia cómo los objetos con clasificación CI se agrupan en el subespacio reducido.



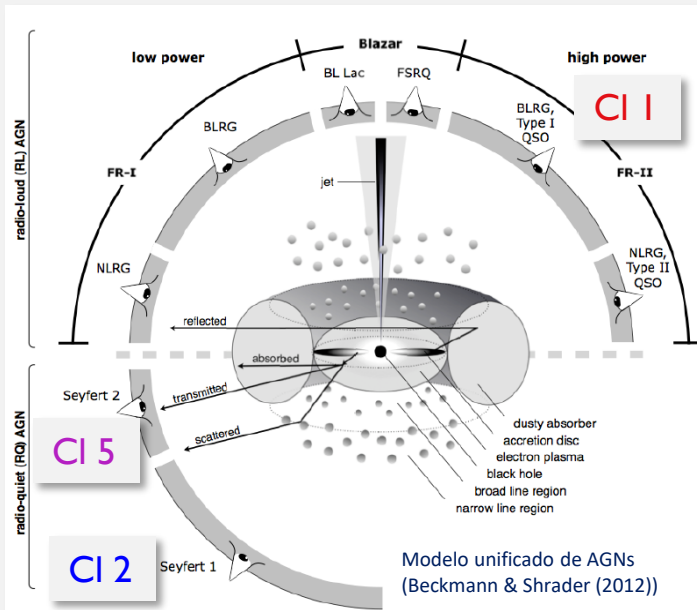
Los **CI 4** son candidatos a cuásar. Se podrían clasificar en un grupo de **CI 1**, **CI 2** o **CI 5**.

Gráfico de dispersión de los puntos del subespacio reducido por ACP a 2 dimensiones



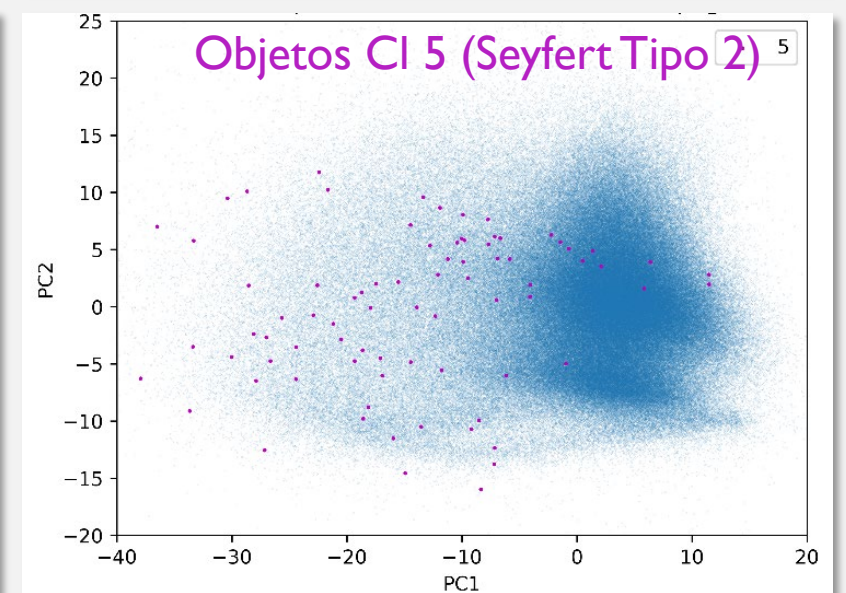
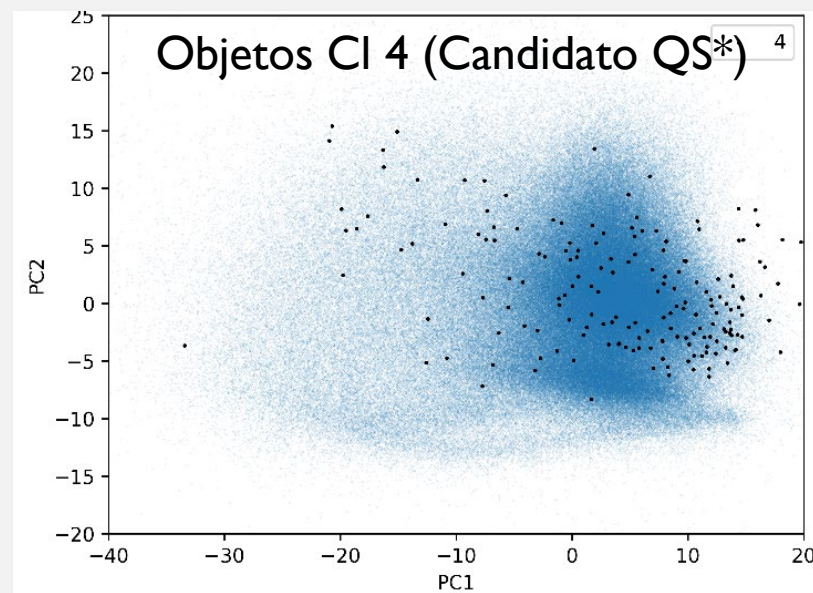
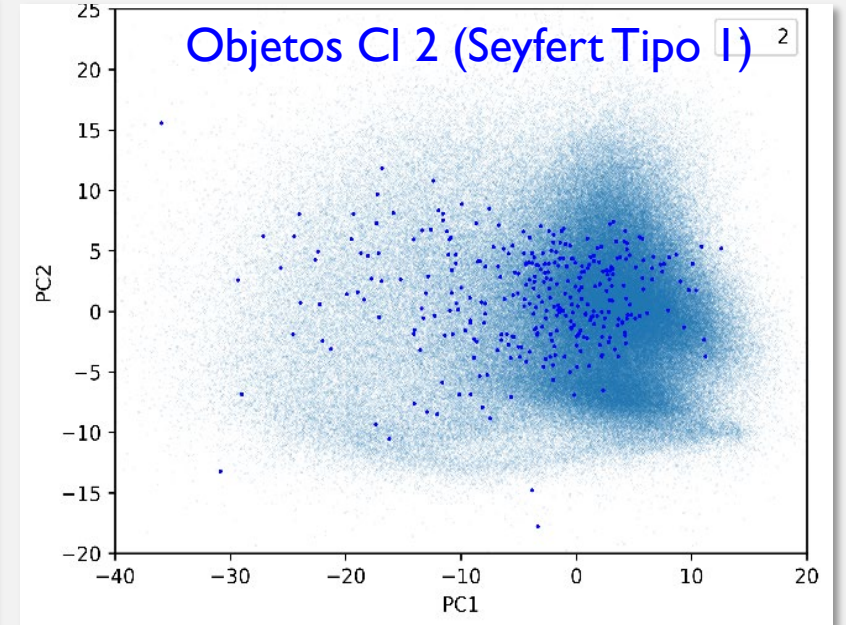
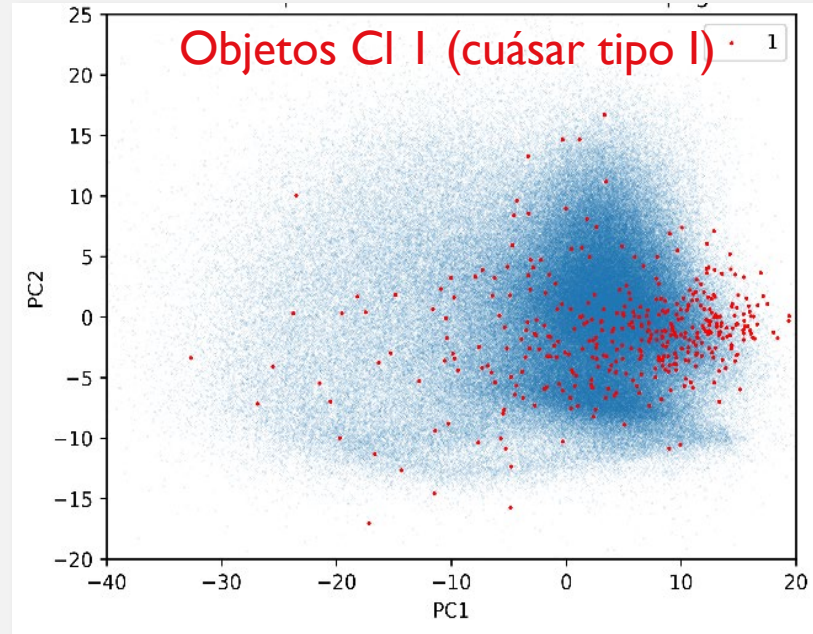
Espacio reducido por PGA

Se aprecia cómo los objetos con clasificación CI se agrupan en el subespacio reducido.



Los CI 4 son candidatos a cuásar. Se podrían clasificar en un grupo de CI 1, CI 2 o CI 5.

Gráfico de dispersión de los puntos del subespacio reducido por PGA a 2 dimensiones



Dispersión color-color comparado con espacio reducido ACP y PGA

Gráfico de dispersión color-color.
Magnitudes lineal.

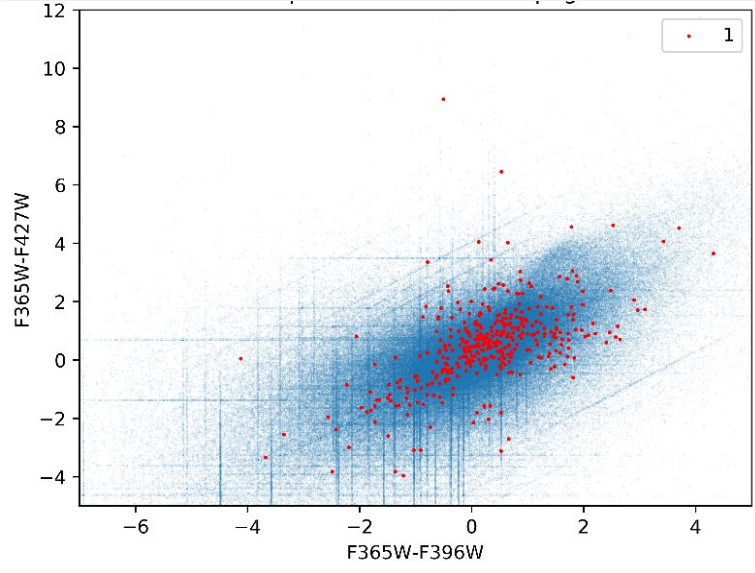


Gráfico de dispersión subespacio ACP
Adimensional.

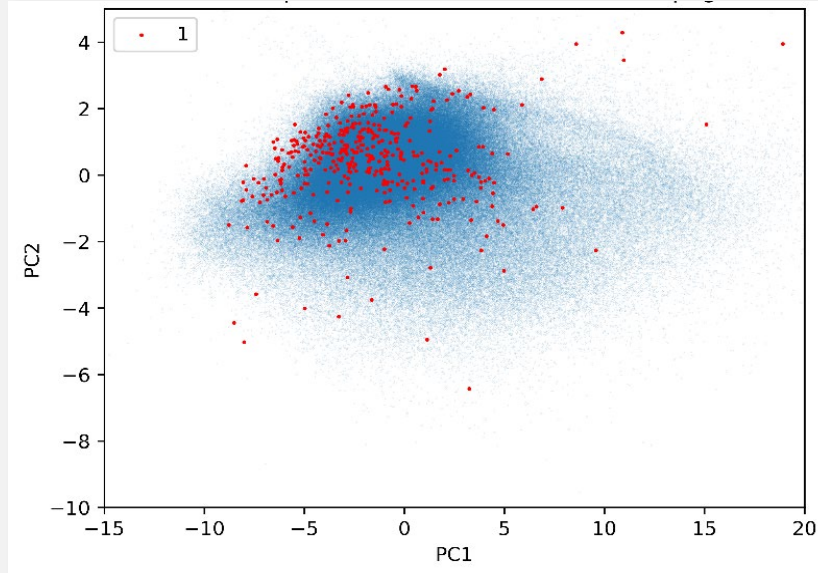
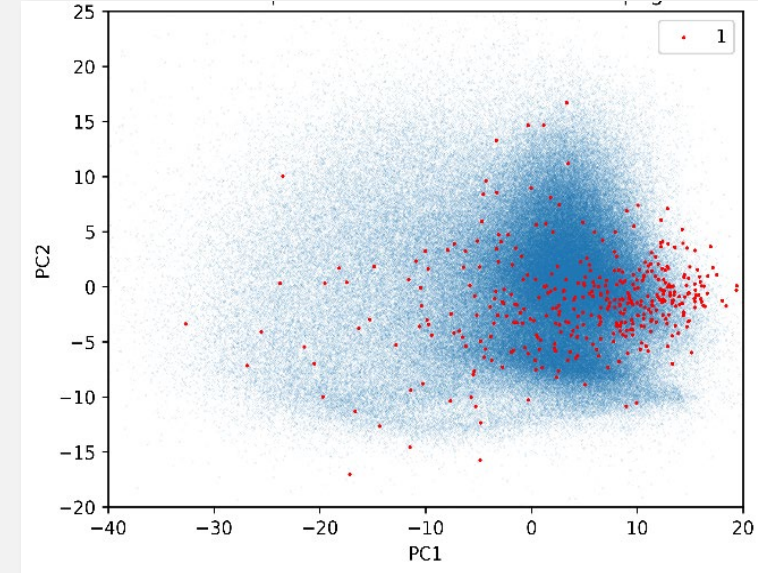
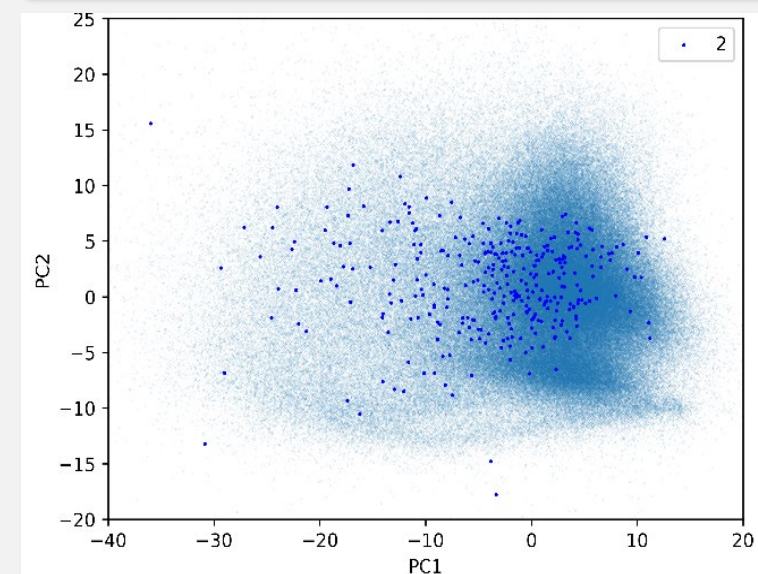
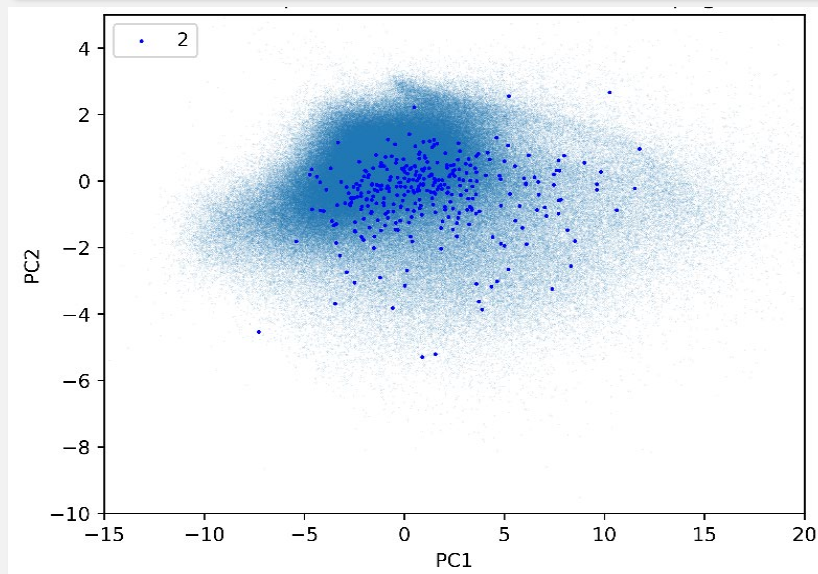
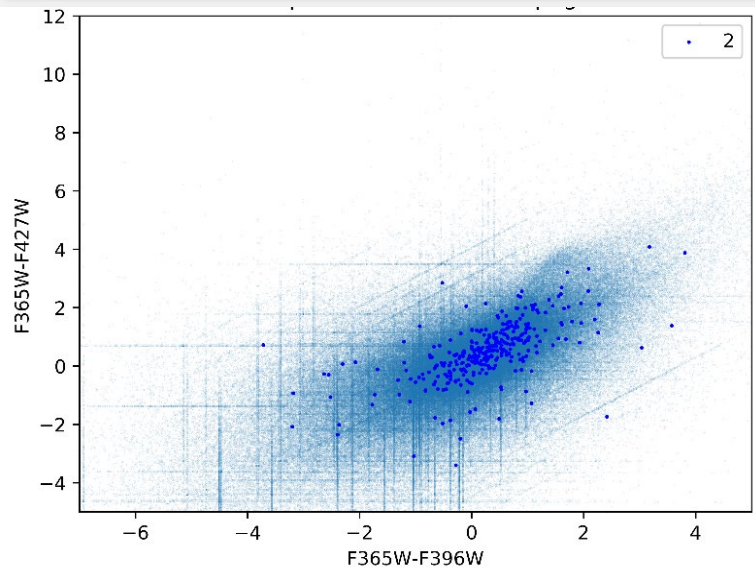


Gráfico de dispersión subespacio PGA
Adimensional.

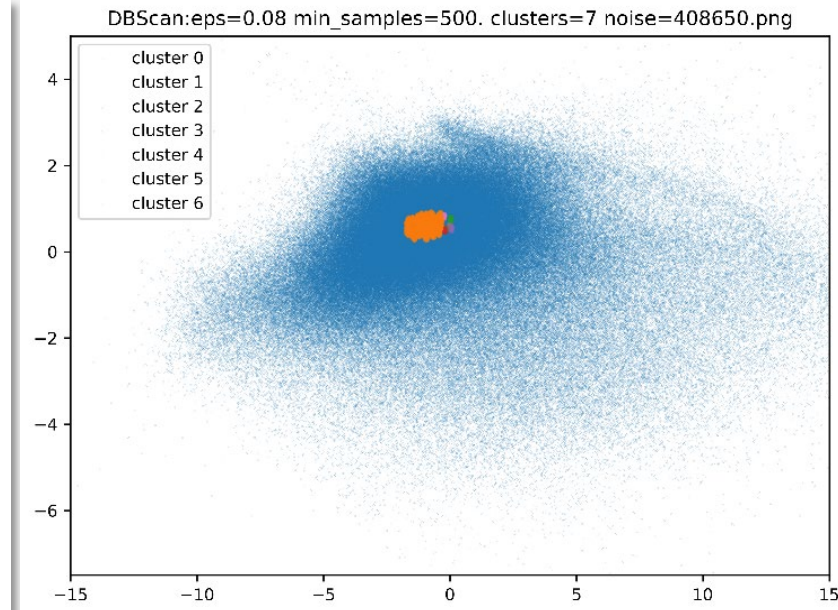
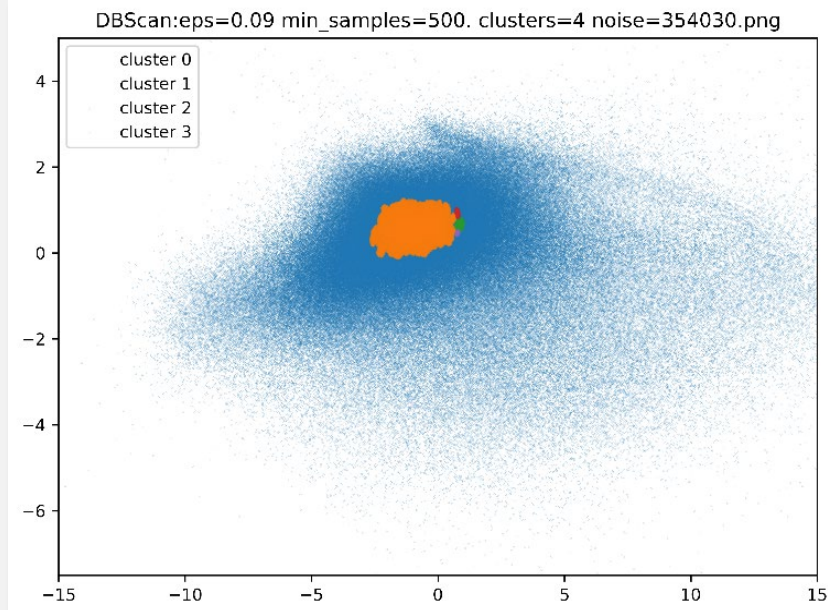
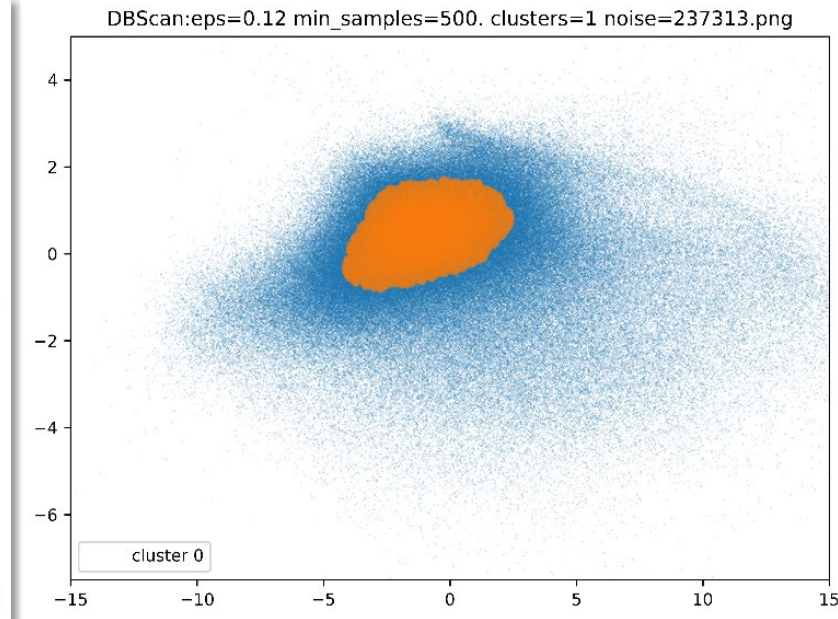
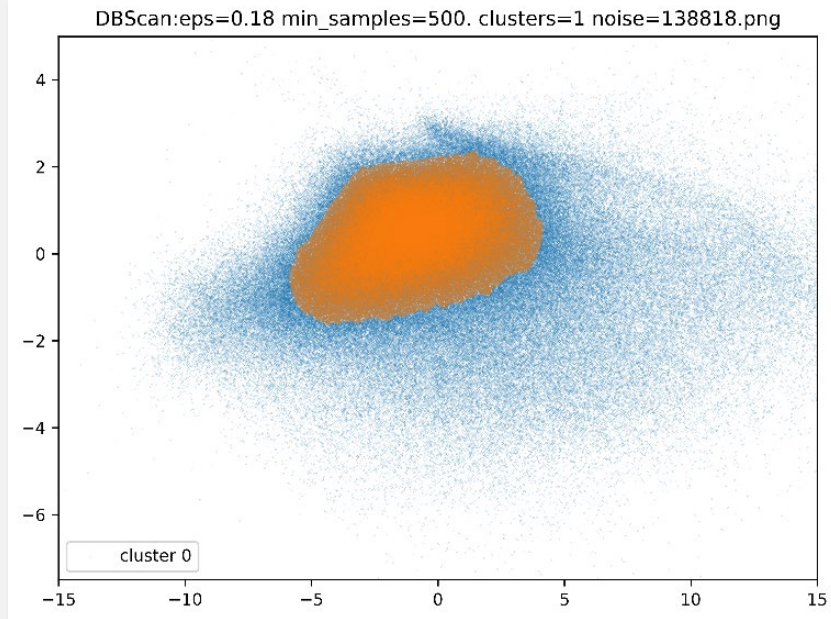


Objetos Cl 1 (cuásar tipo I)

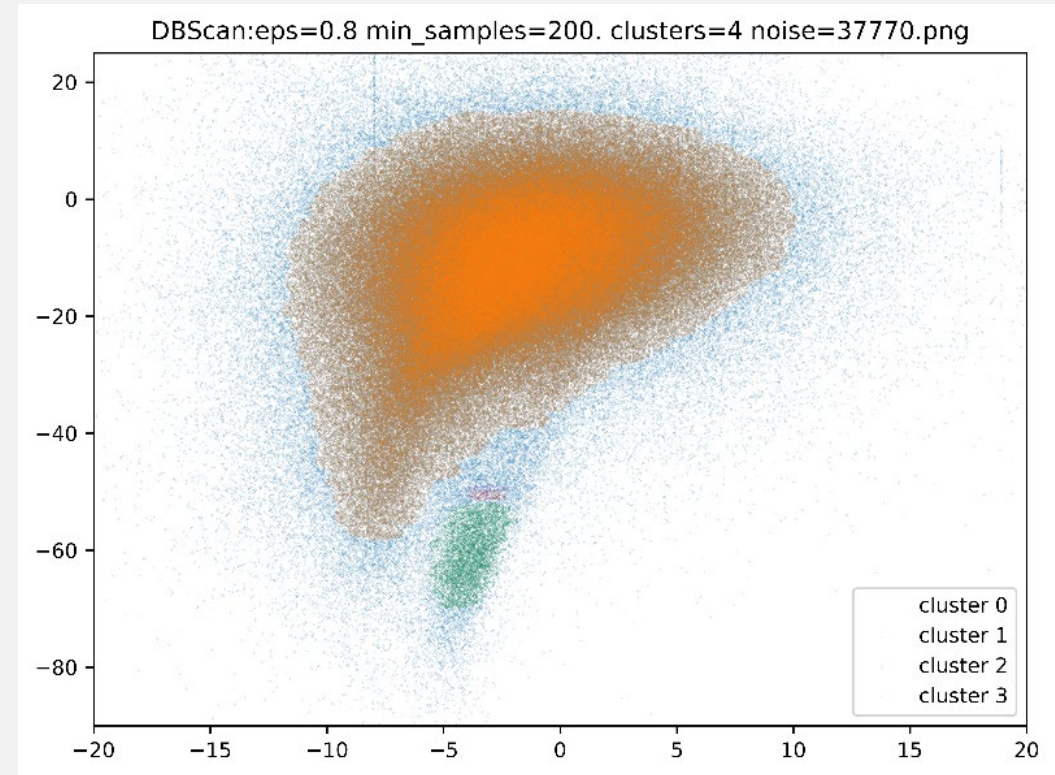
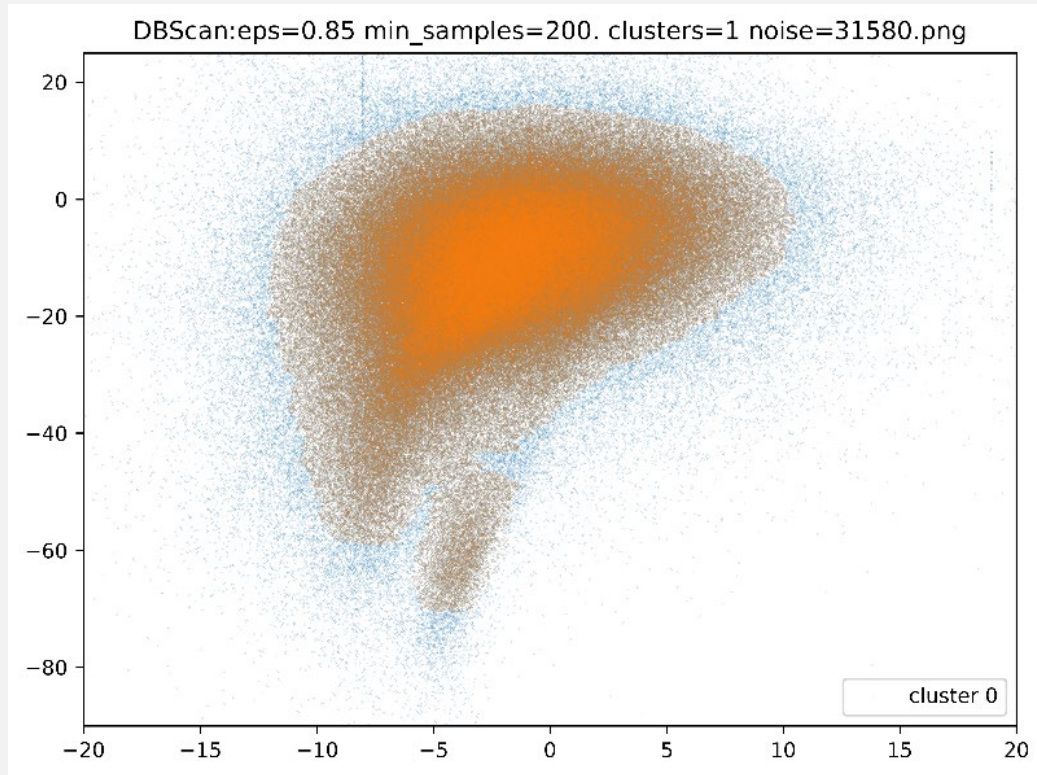
Objetos Cl 2 (Seyfert Tipo I)



Algoritmos de clusterización – DBScan sobre ACP



Algoritmos de clusterización – DBScan sobre SRP



conclusiones

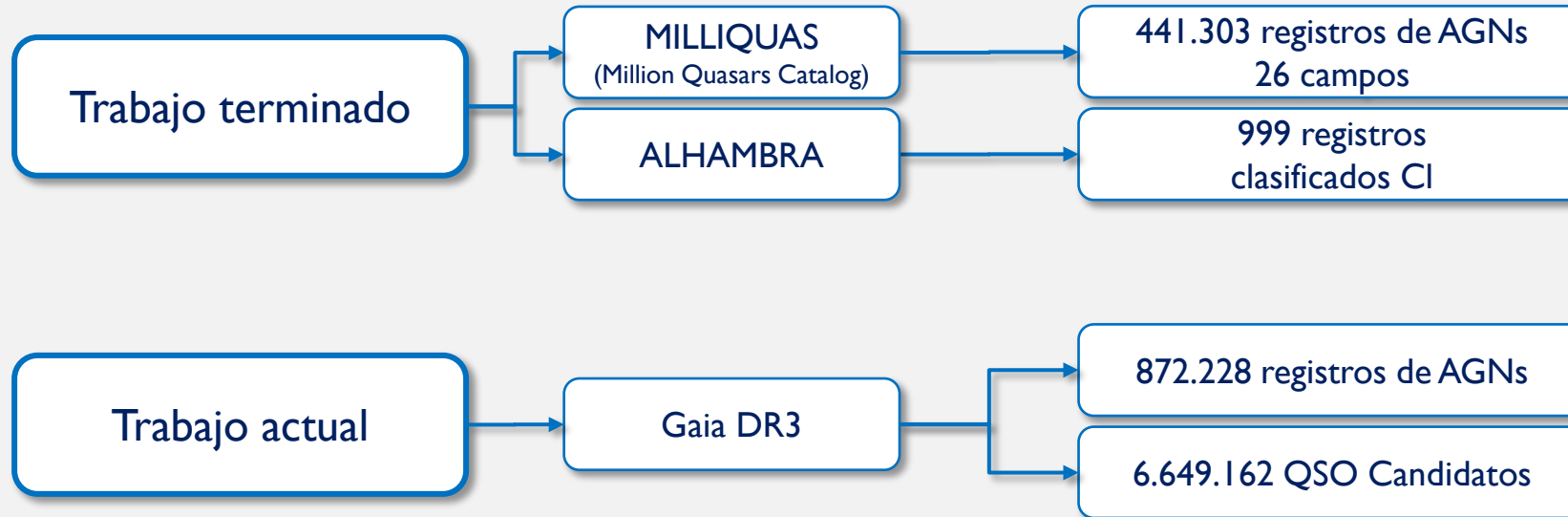
Analizar los resultados respecto de los modelos actuales

Descubrimiento de objetos no esperados o exóticos

Hay mucha información debajo de la enorme cantidad de datos en los catálogos

Cada vez se generan más datos a la espera de ser analizados

nuevas líneas de trabajo



The logo for the XXV Congreso Estatal de Astronomía, consisting of the letters 'CEEA' in a stylized, white, sans-serif font. The 'C' is a simple circle, the 'E' has three horizontal bars, and the 'A' has a dot on its right vertical stroke.

CEEA

XXV CONGRESO ESTATAL
DE ASTRONOMÍA

ZARAGOZA
28 - 30 abril 2023

Aprendizaje automático en catálogos astronómicos

Esteban Fernández Mañanes

