

## Identificación de los ficheros de las muestras:

Hay tres formatos:

1. G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_muestra\_tot\_meA.csv
2. G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_muestra\_70\_tot\_meA.csv
3. G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_galaxias\_intA.csv

**Formato 1.** La muestra de entrenamiento contiene el 60% de las observaciones y la de prueba el 40% restante. Contienen coordenadas (x, y, z) de celdas donde la unidad es el ancho de celda, además, se indica si la celda está ocupada (columna E) y si es así por cuantas galaxias (columna n).

G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_muestra\_tot\_meA.csv

### G\_1

G\_0 distancias de luminosidad

G\_1 distancias comoviles

### \_10

Diámetro del universo en Gal

### 444

Número de celdas por eje: X, Y, Z

Delimitación de la corona esférica:

### 8\_16\_

Ente 6 y 14 horas de ascensión recta

### 15\_70

Entre 15 y 70 grados de declinación

### 600000\_1000000

Entre 600 y 1000 millones de años luz respecto a la tierra

muestra\_tot\_meA

muestra\_tot\_meA = muestra de entrenamiento hyperleda

muestra\_tot\_mpA = muestra de prueba hyperleda

muestra\_tot\_mzA = agregado de las dos anteriores

muestra\_tot\_az\_meA = muestra de entrenamiento generada al azar

muestra\_tot\_az\_mpA = muestra de prueba generada al azar

muestra\_tot\_az\_mzA = agregado de las dos anteriores

muestra\_tot\_azaj\_meA = muestra de entrenamiento generada al azar ajustada por la distancia

muestra\_tot\_azaj\_mpA = muestra de prueba generada al azar ajustada por la distancia

muestra\_tot\_azaj\_mzA = agregado de las dos anteriores

**Formato 2.** Es igual que el anterior salvo que añade el porcentaje de observaciones en la muestra de entrenamiento:

G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_muestra\_70\_tot\_meA.csv

**\_70\_**

La muestra de entrenamiento contiene el 70% de las observaciones.

**Formato 3.** Similar a los anteriores.

G\_1\_10\_444\_8\_16\_15\_70\_600000\_1000000\_galaxias\_intA.csv

Contiene tantas observaciones como galaxias hay en el sector. Para cada observación (galaxia) registra las coordenadas de la celda en la que está incluida, todas las observaciones contienen un 1 en la columna de ocupación (columna E) y un 0 o un 1 el que corresponde al número de galaxias de los formatos 1 y 2 (columna n), el 1 significa que la galaxia pertenece a la muestra de entrenamiento y el 0 a la de prueba. También puede contener otros datos (objname;pgc;al2000;de2000;r)