Búsqueda de estructuras en "datos puntuales"

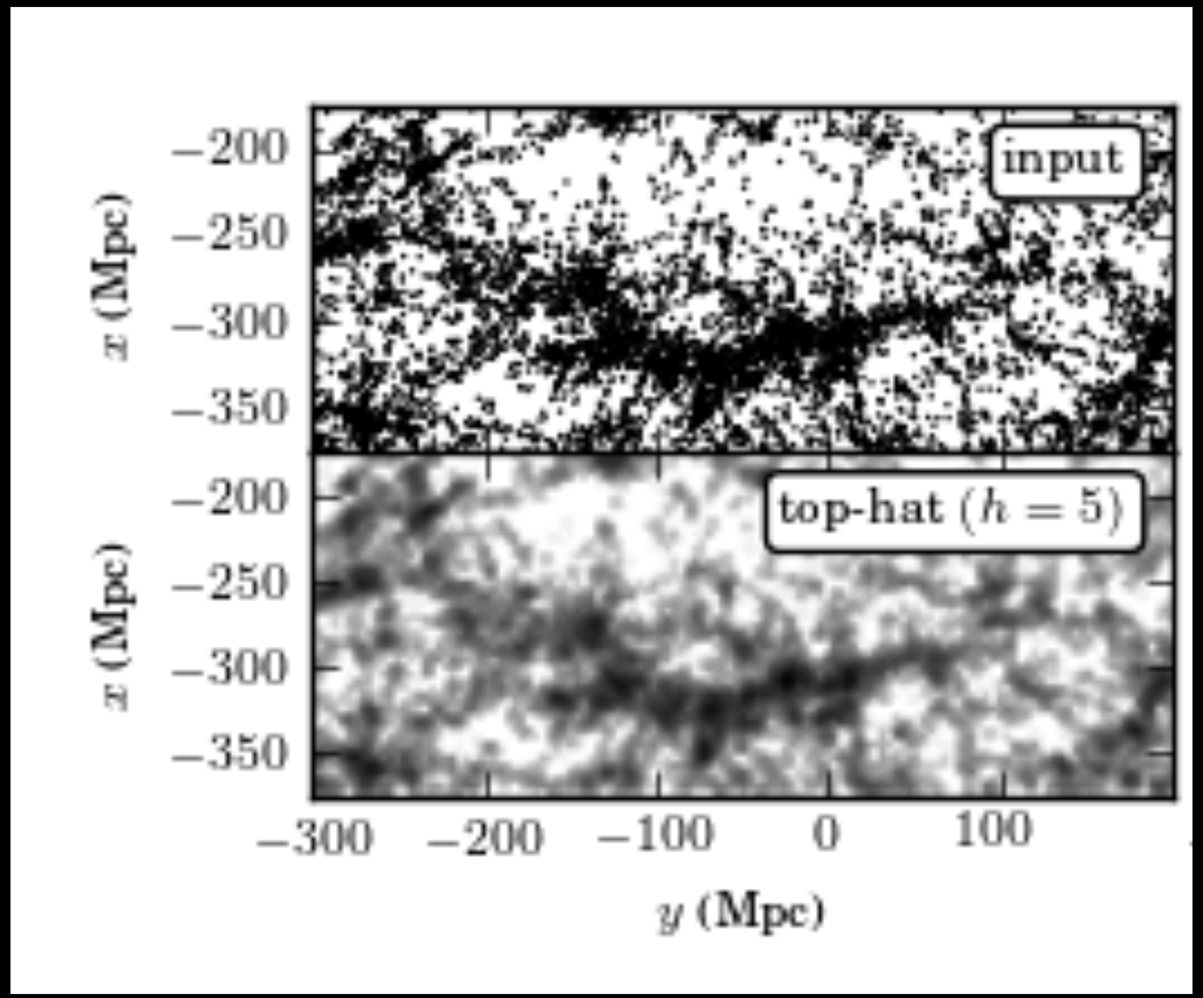
- Datos Puntuales: cantidad que pueda ser representada en un espacio coordenado.
  - Puede ser que por la naturaleza del conjunto de datos dicho espacio sea precisamente un espacio físico, por ejemplo las posiciones y velocidades de un conjunto de puntos son representadas en un espacio 6-dimensional.
  - Dichos puntos pueden representar desde partículas fundamentales, objetos macroscopicos, hasta galaxias en el Universo.
  - Los puntos pueden representar atributos, o mediciones sobre un sistema; por ejemplo la distribución de autos de cierto tipo en una ciudad; la distribución de personas infectadas en una población, etc.

Busqueda de estructuras en datos puntuales

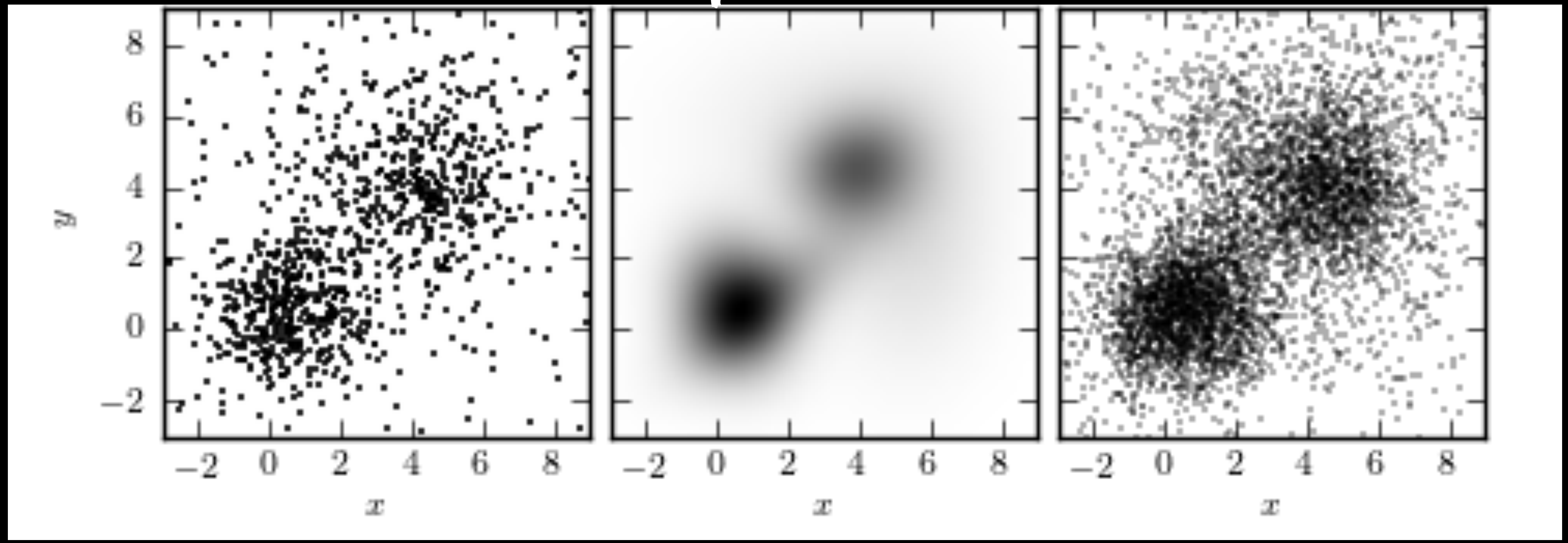
- Dada una muestra de N puntos en un espacio D-dimensional hay tres posibles técnicas a explorar:
    - Estimación de densidad ---> Inferencia de una distribución de probabilidad
    - Búsqueda de agrupamientos (clustering) ---> búsqueda de estructuras
    - Descripción estadística de la estructura observada.

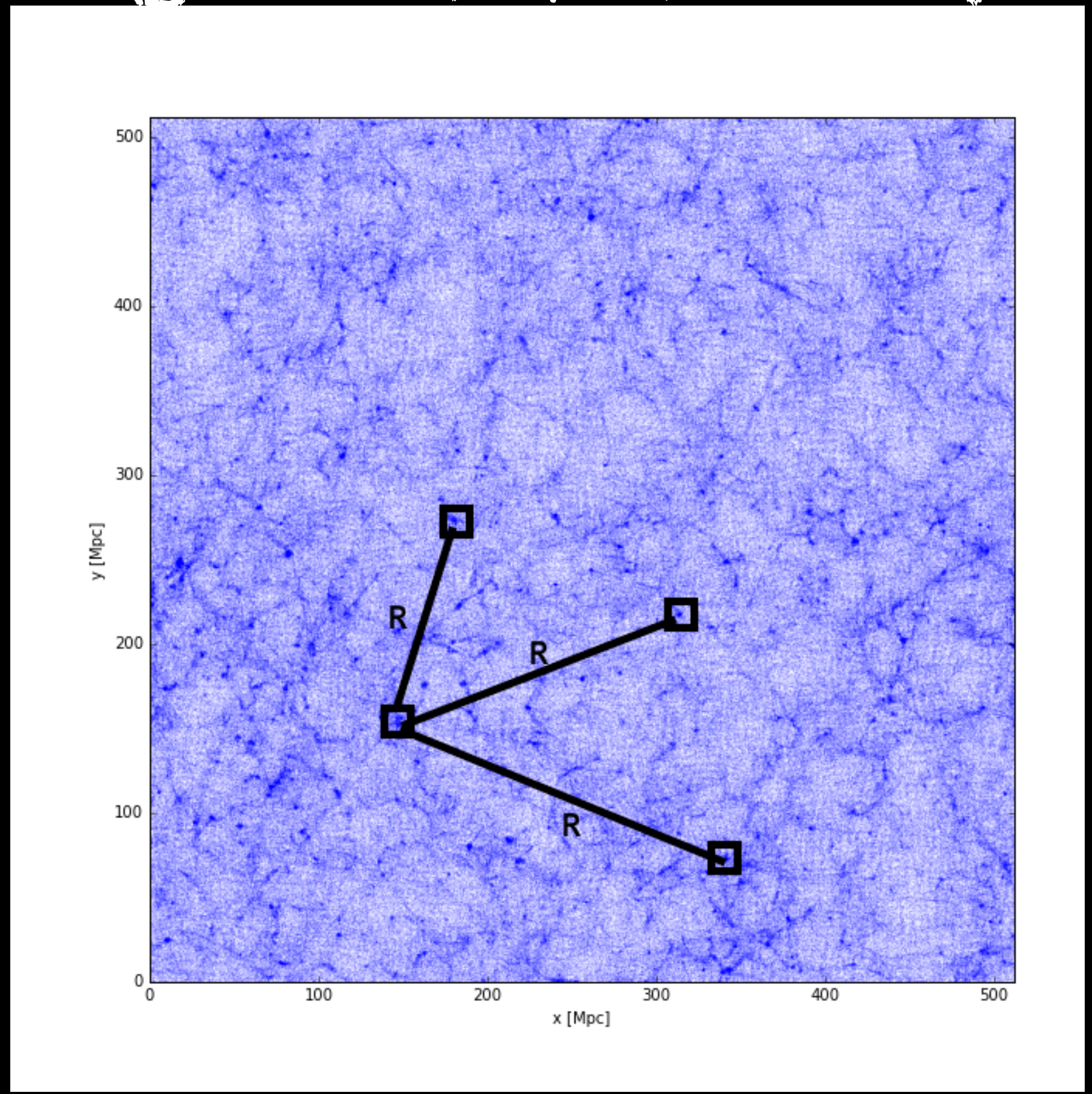
¿distribución?

Densidad

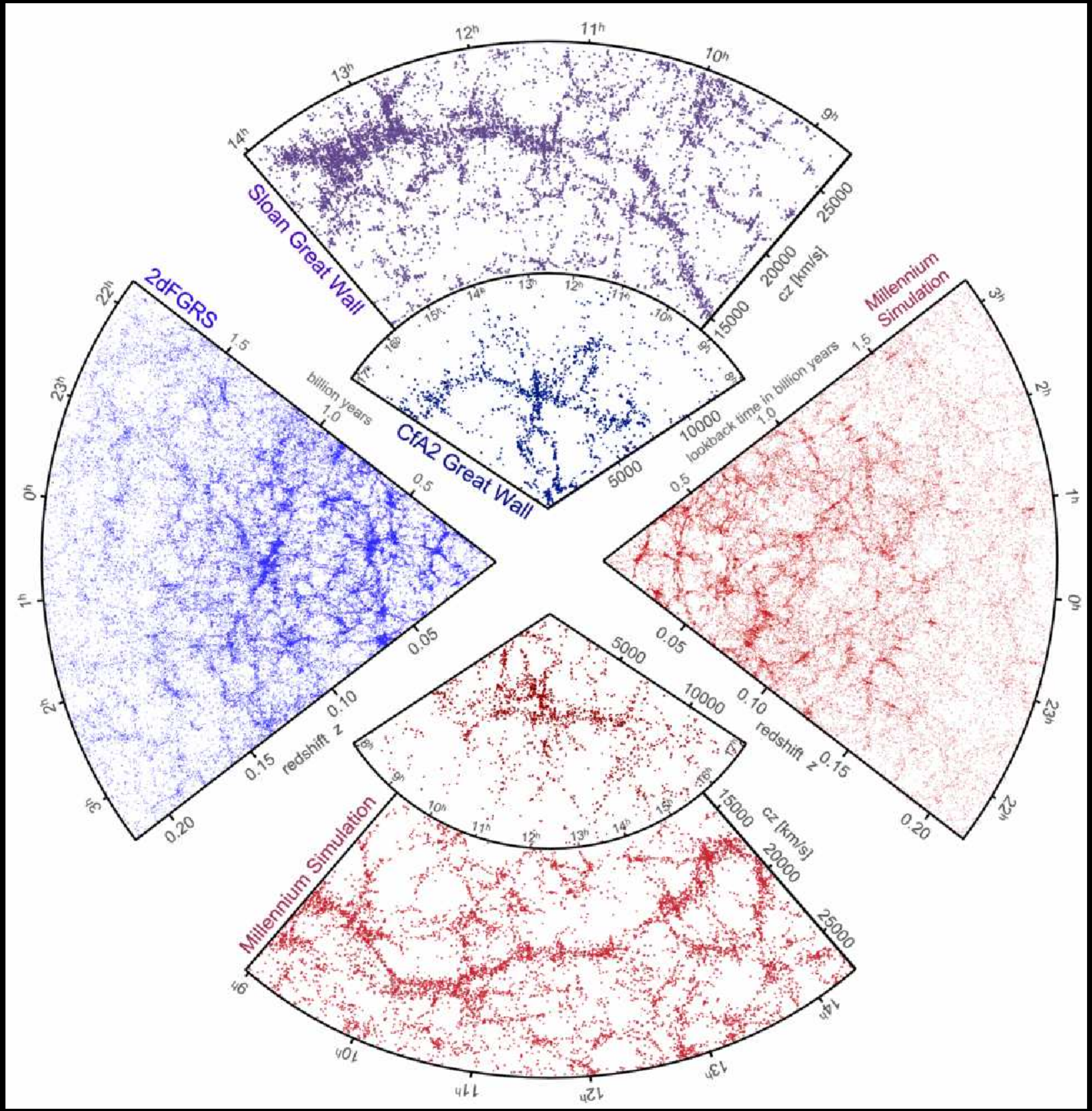


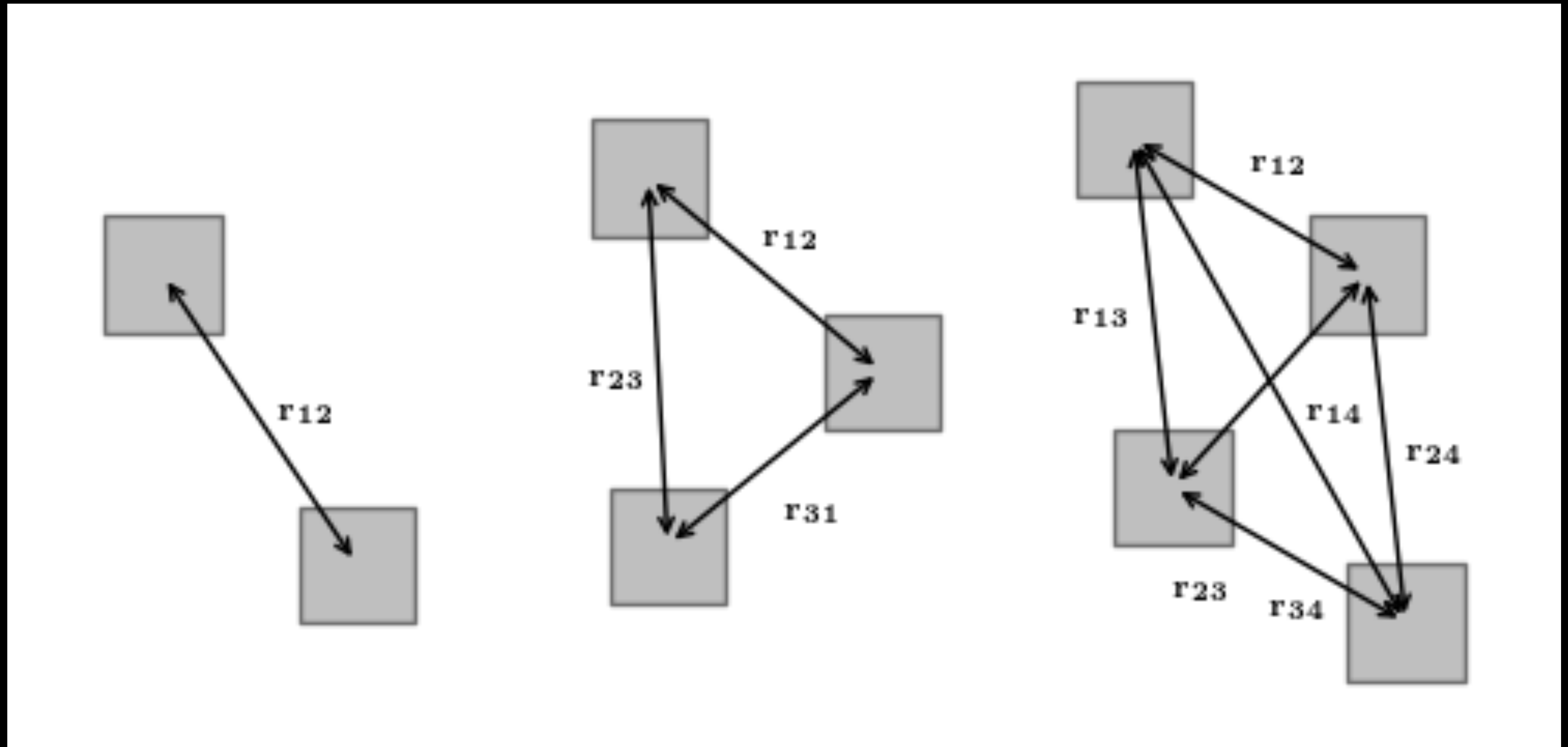
¿agrupamientos?





Descripción estadística: Funciones de Correlación N Puntos

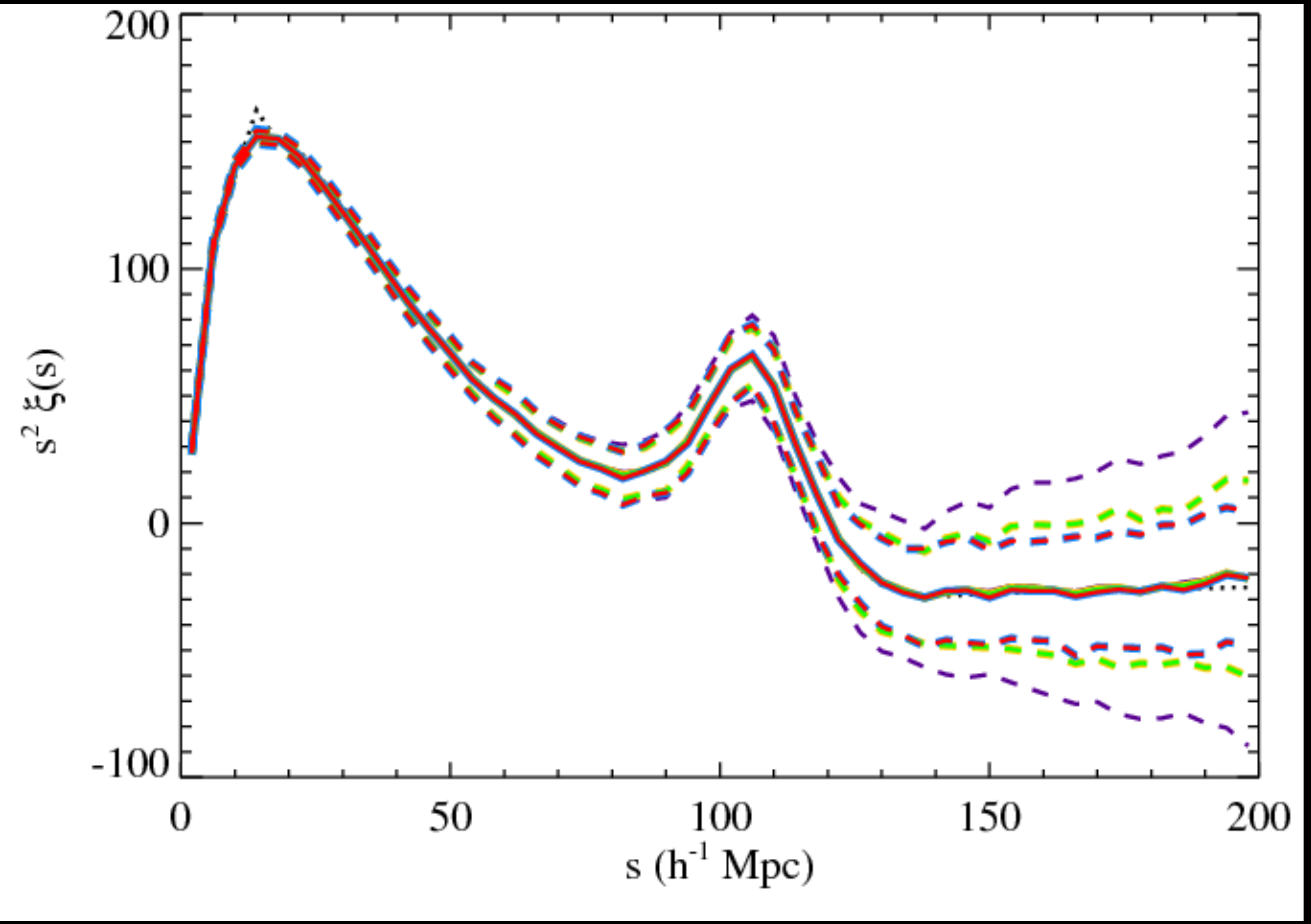


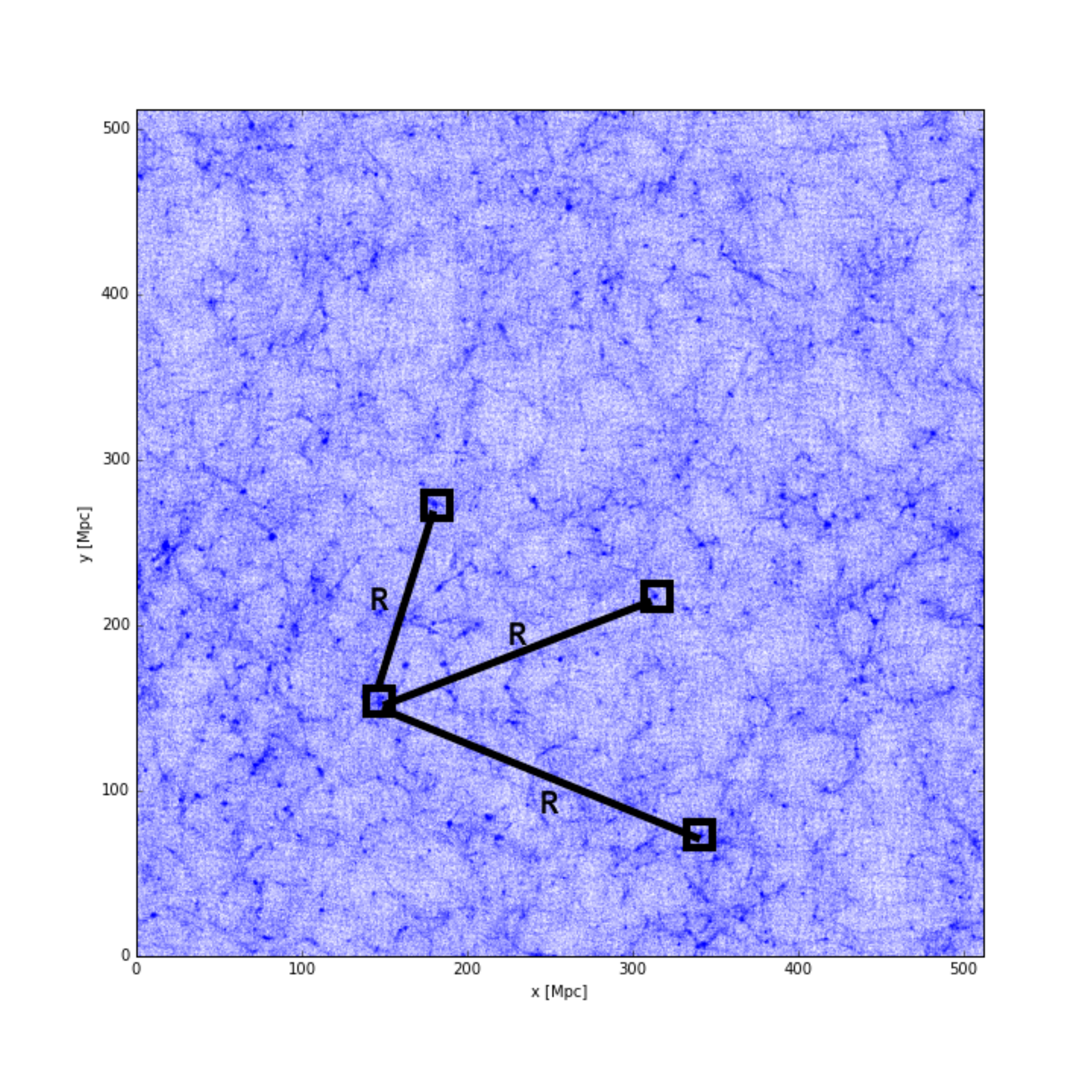


(a) 2 point (b) 3 point (c) 4 point

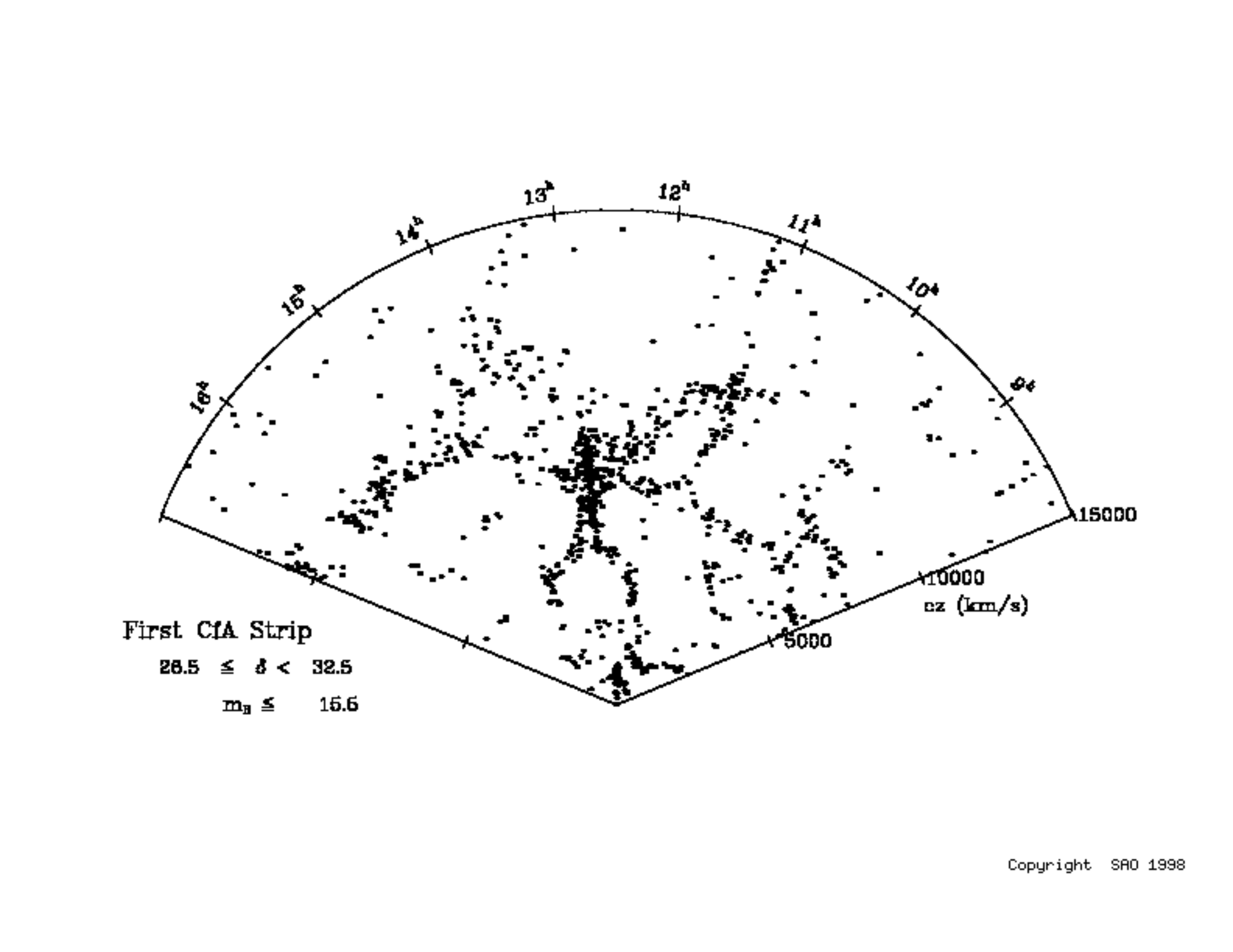
¿Qué tanto difiere la distribución de una medida aleatoria?

Función de correlación de 2-puntos (2PCF)



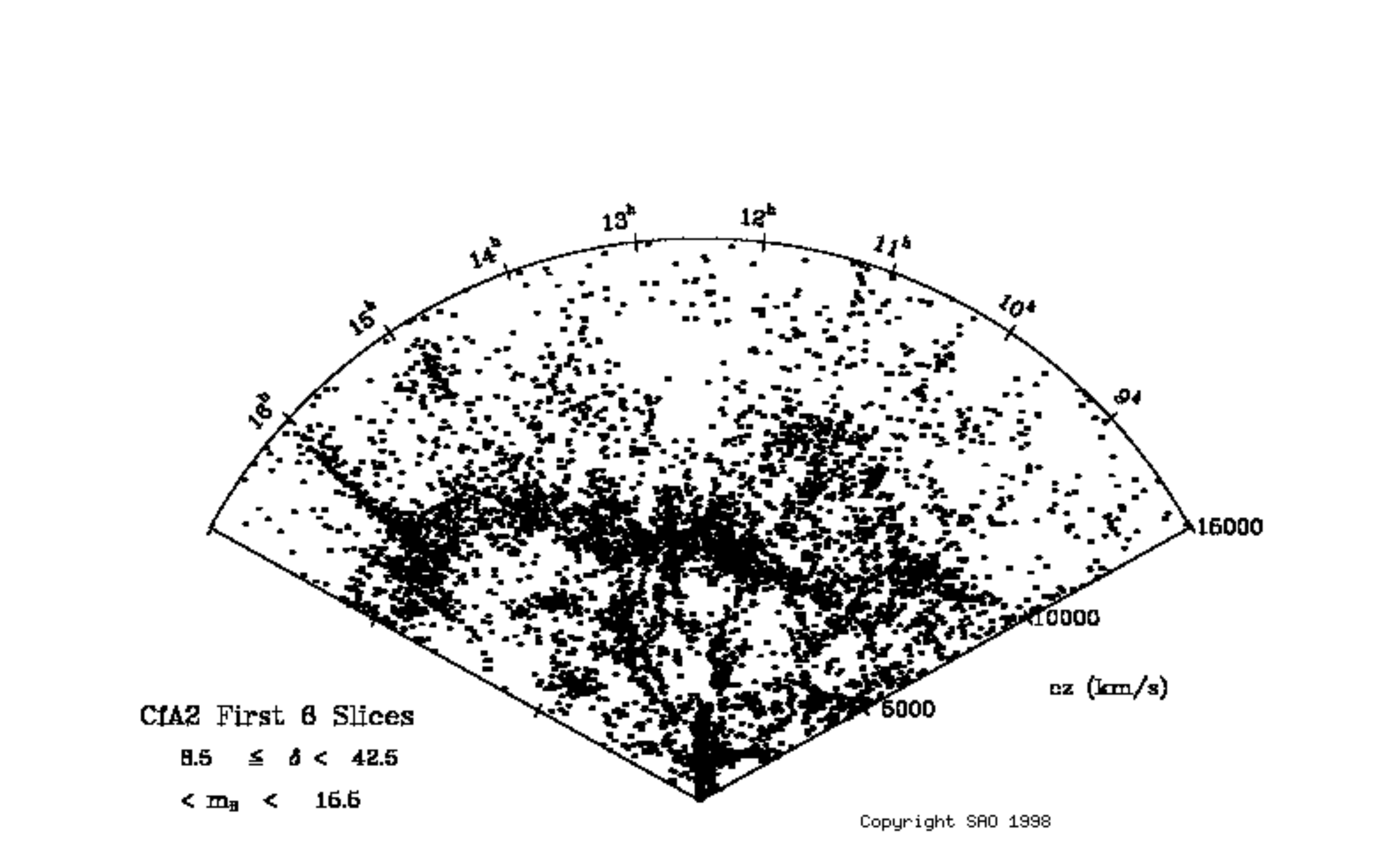


Censos de Galaxias (2PCF)

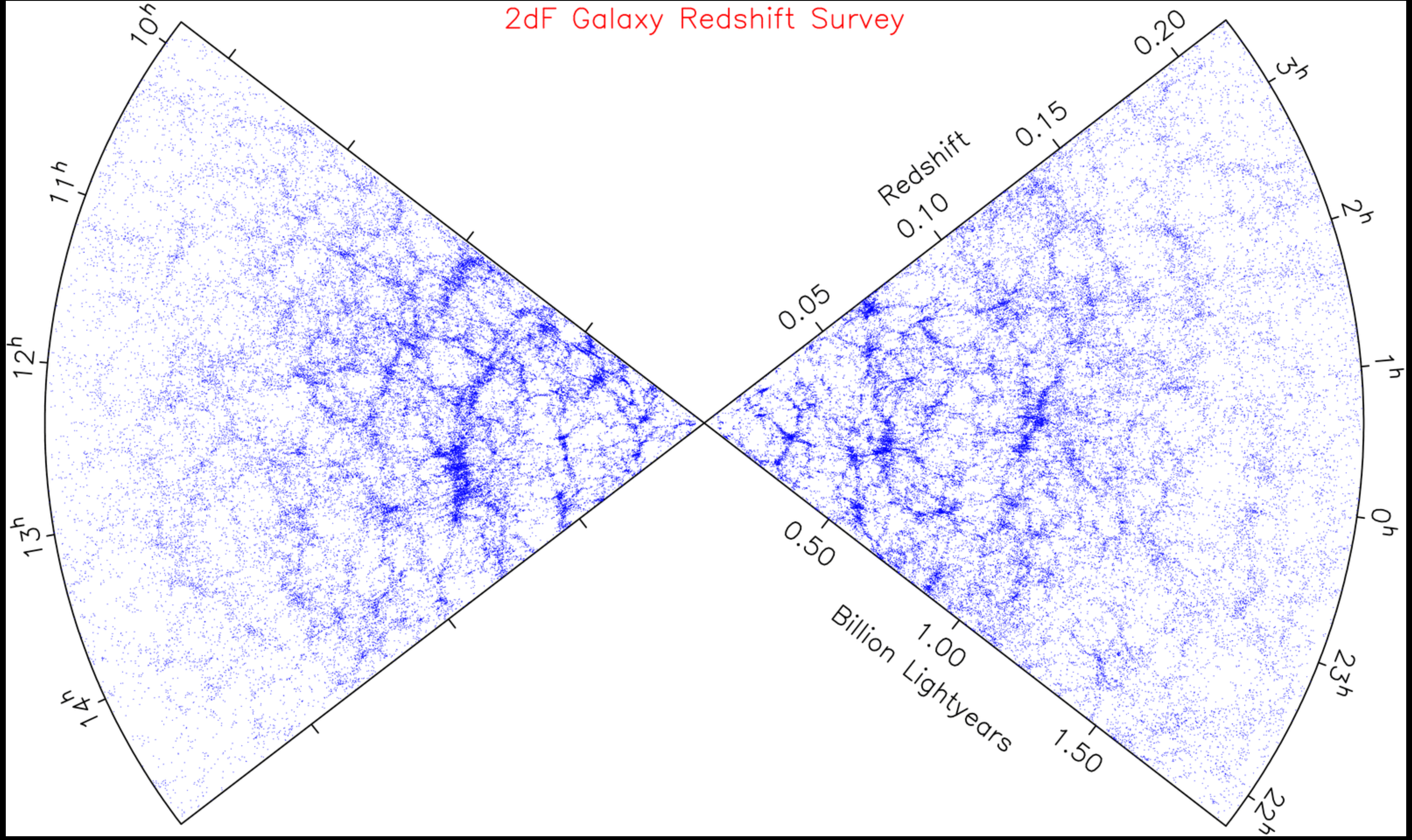


1985:
~1110 Galaxias

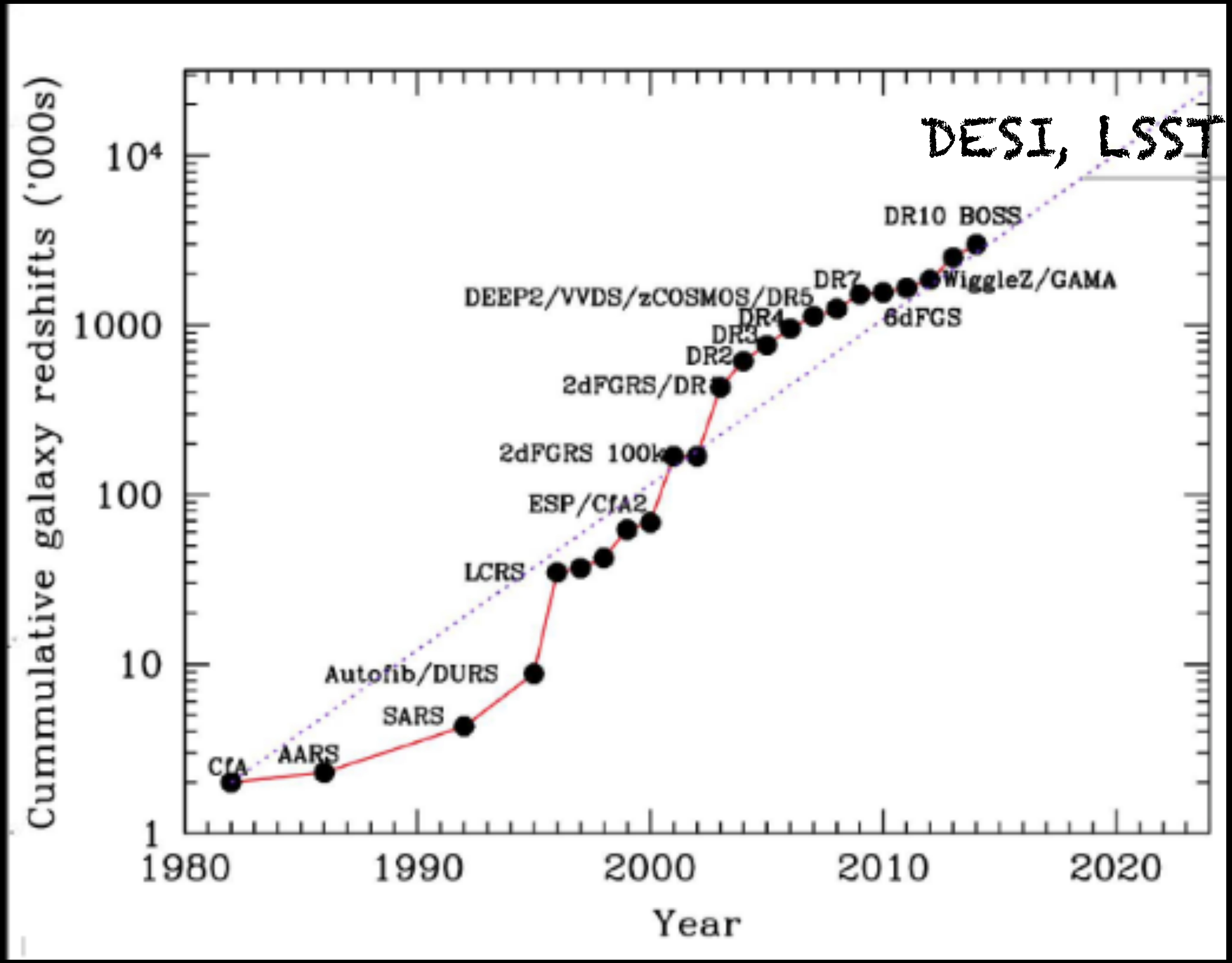
1989:
~14000 Galaxias



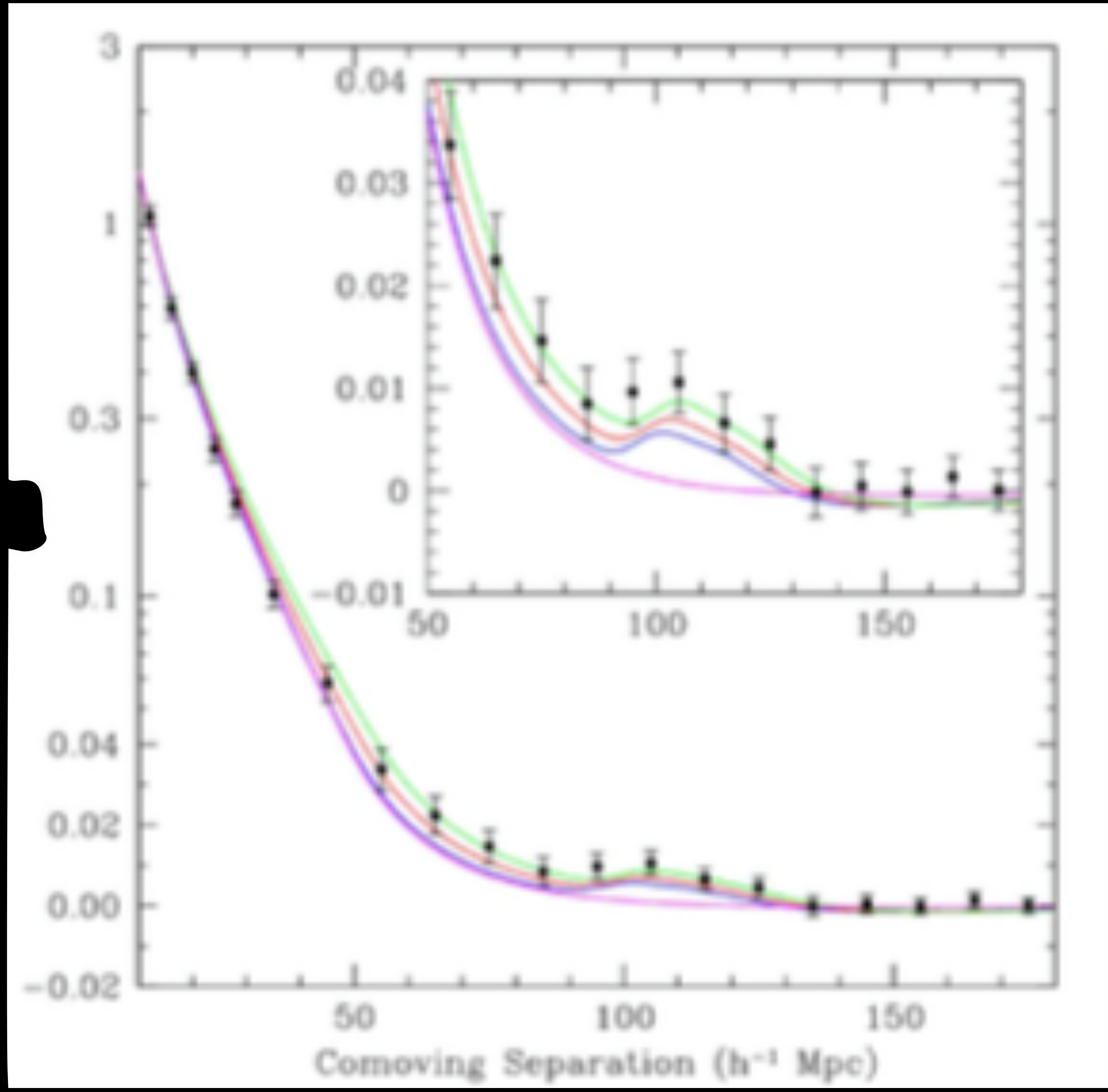
Censos de Galaxias (2PCF)

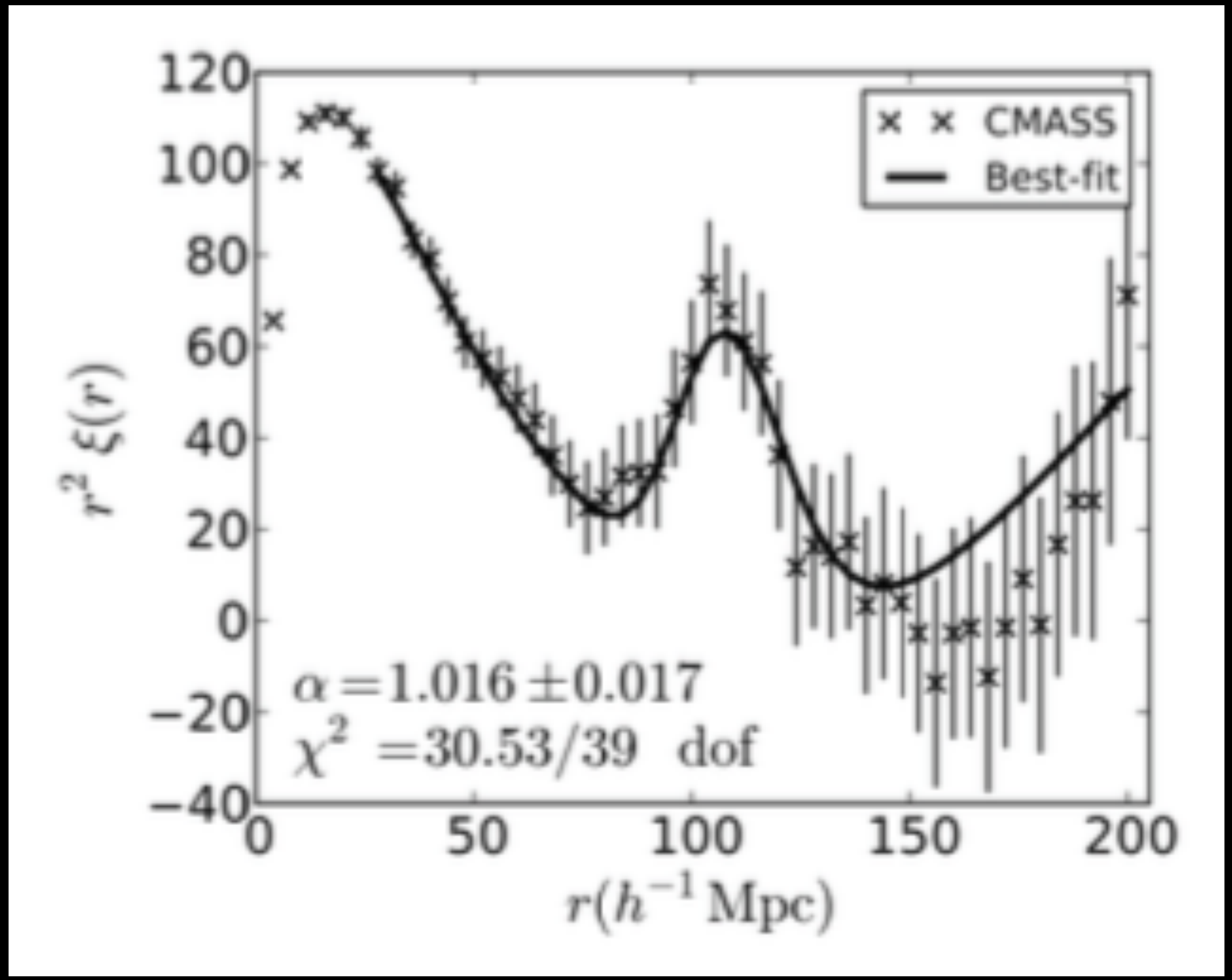


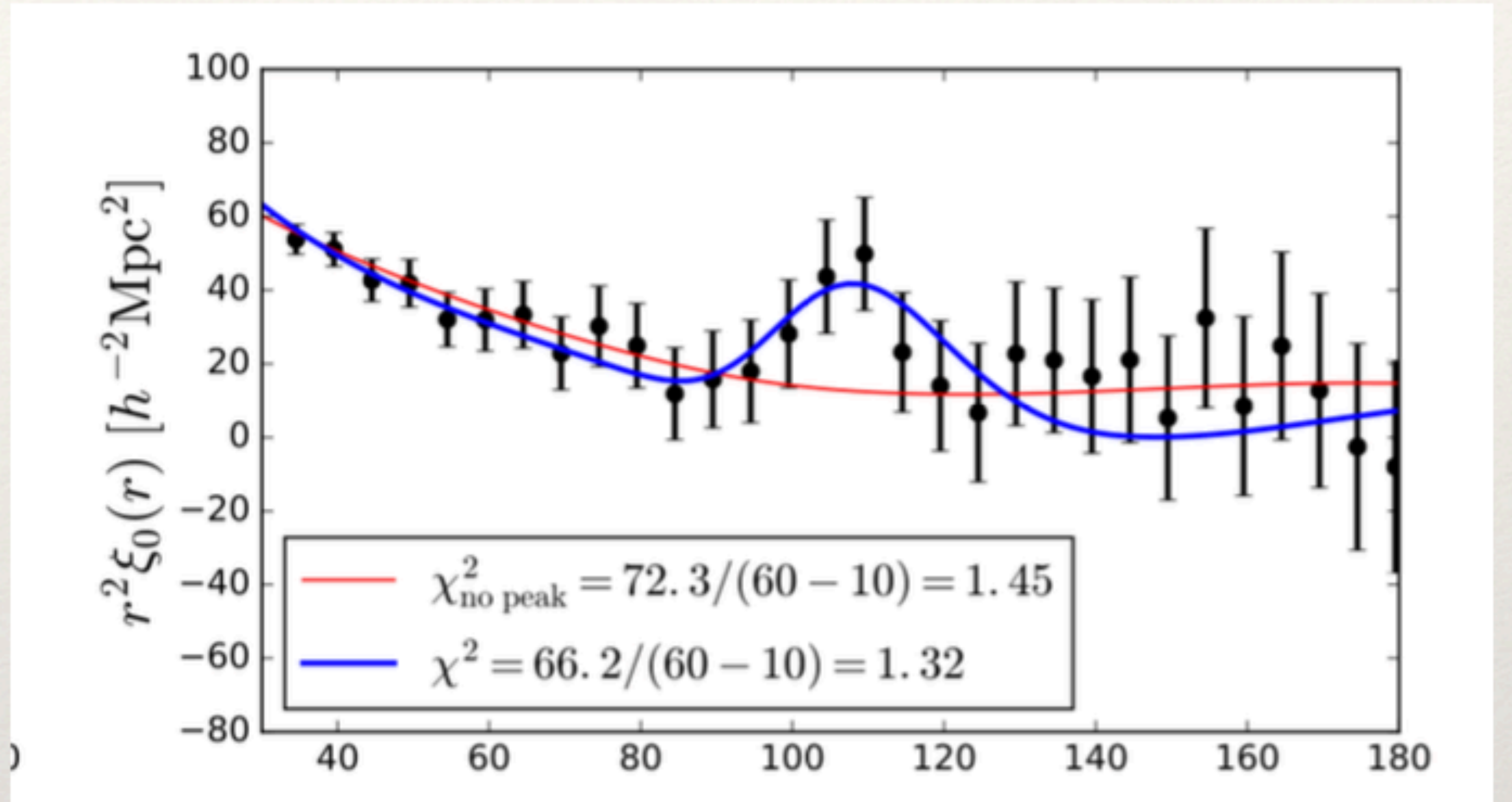
~220,000 Galaxias.
2005



Sondeos de Galaxias (2PCF)







Bautista, M. Vargas, et. al 2018

Función de correlación de 2-puntos (2PCF)

$$dP_{12} = \rho^2 dV_1 dV_2 (1 + \xi(r))$$

P_{12} Probabilidad de encontrar un par de puntos en dos elementos de Volumen separados a una distancia r.

$\xi(r)$ Función de correlación: Exceso de probabilidad de encontrar un par de puntos, comparado con una distribución aleatoria.

Positiva, Negativa, o cero ——-> distribución correlacionda, anticorrelacionada, aleatoria

$$\xi(r) = \frac{1}{2\pi^2} \int dk k^2 P(x) \frac{sin(kr)}{kr}$$

Transformada de Fourier del espectro de potencias.

$$\xi(r) = < \frac{\delta\rho(x)}{\rho} \frac{\delta\rho(x+r)}{\rho} >$$

$$\delta\rho(x)/\rho = (\rho(x) - \bar{\rho})/\rho(x)$$

Estimadores mas simples de 2PCF

$$\xi(r) = \frac{n_D DD(r)}{n_R RR(r)} - 1$$

$$\xi(r) = \frac{DD(r) - 2DR(r) + RR(r)}{RR(r)}$$

Landy-Szalay

$DD(r)$: Número de pares de puntos separados a una distancia r en la distribución de datos

$RR(r)$: Número de pares de puntos separados a una distancia r en una distribución aleatoria

$DR(r)$: Número de pares de puntos separados a una distancia r, usando la distribución de datos y la aleatoria como una sola.

r: en es el intervalo en el que se está calculando el número de pares.

Ejercicio

- Genera una distribución de puntos aleatorias (Uniformes) en un una cubo (o caja) de lado L.
- Haz un programa que cálcele la función de correlación y verifica que en el cubo generado, o en otro cubo generado también aleatorio, ésta es cero. (Permite que tu código trabaje en 2 y 3 dimensiones)
- *Use el código para calcular la 2PCF en el conjunto de datos que te voy a proporcionar.
- ¿Cómo harías más eficiente el cálculo?