

A. FIGURAS DE CLÚSTERES Y ÁRBOLES

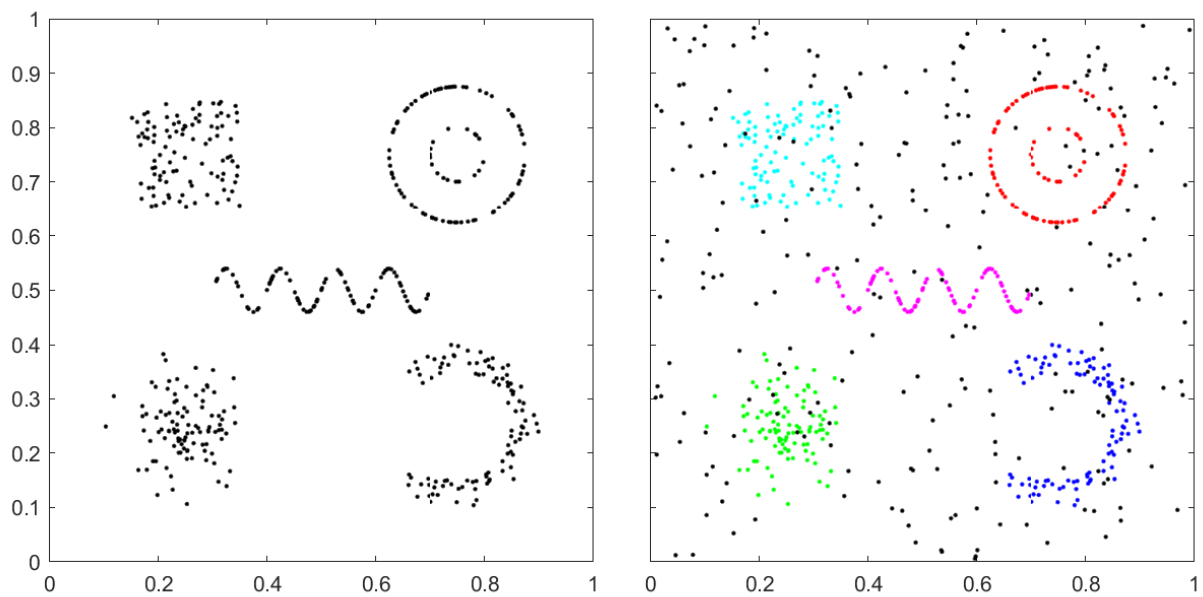


Fig. A.1: Benchmark creado para visualizar los algoritmos de clusterización (figura izquierda). Las características de los clústeres son: delimitado con densidad constante (arriba izquierda), lineal contenido dentro de otro (arriba derecha), lineal alargado (centro), distribución normal (abajo izquierda) y cóncavo con densidad constante (abajo derecha). Cada elemento contiene 100 puntos. Sobre estos puntos se han superpuesto 250 puntos de ruido distribuidos uniformemente (figura derecha).

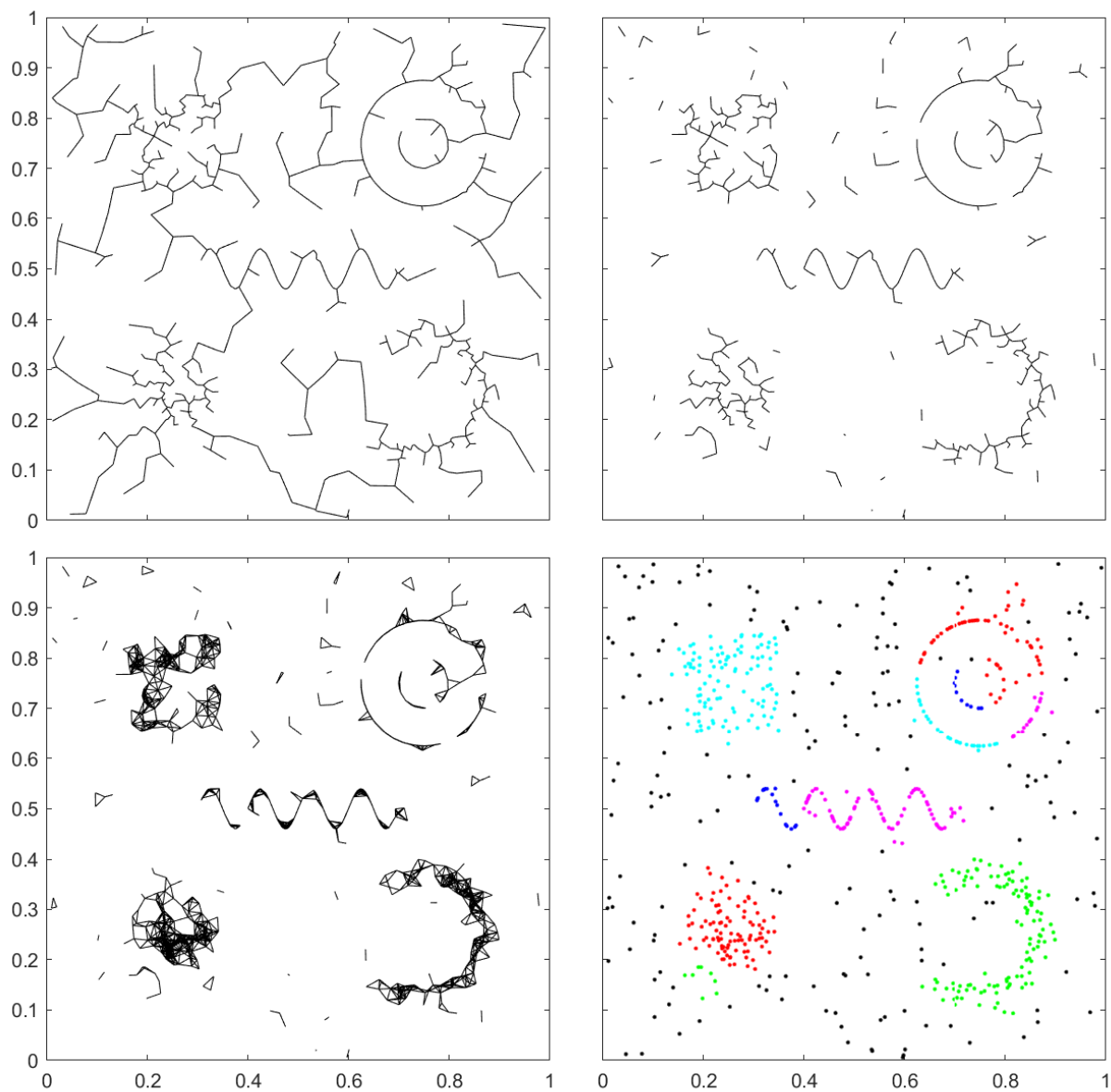


Fig. A.2: Visualización de los pasos de las distintas versiones del algoritmo MST. El parámetro de distancia máxima es $b = 0.03$. Para que un subgrafo se considere un clúster necesita contener al menos 5 nodos. Se muestra primero el MST completo generado con la versión **MST_v1** (arriba izquierda) y el árbol después de eliminar las aristas con una distancia $> b$ (arriba derecha). Se muestran también los subgrafos obtenidos después de eliminar las aristas de distancia $> b$ del grafo completo (abajo izquierda). Se muestran con distintos colores los 10 clústeres obtenidos, siendo el resultado el mismo usando cualquiera de las 3 versiones (abajo derecha). Con el algoritmo MST es difícil delimitar claramente los clústeres. Además, dado que la distancia máxima es fija, puede ocurrir que un clúster se separe en dos, o que dos clústeres próximos se unan.

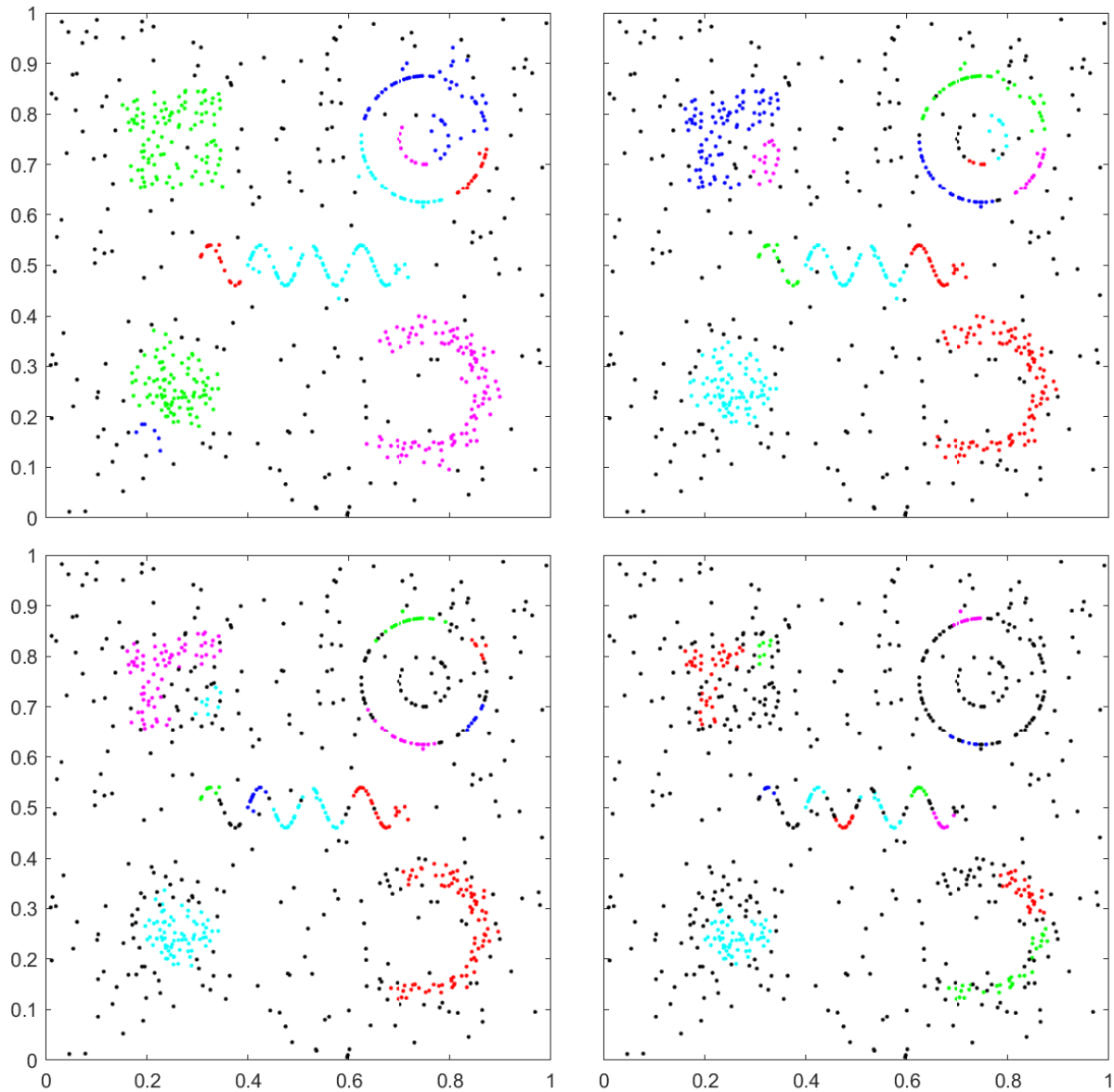


Fig. A.3: Clusterización con DBSCAN para distintos número mínimo de vecinos con misma distancia máxima de enlace $b = 0.03$ y tamaño mínimo de clúster $nMin = 5$. Los distintos valores de $minVecinos$ que se han usado son: 1 (arriba izquierda), 3 (arriba derecha), 5 (abajo izquierda) y 7 (abajo derecha). Para $minVecinos = 1$ los resultados son idénticos a los obtenidos con MST y DTfOf. Según se aumenta el número de vecinos necesarios para que un punto forme parte de un clúster, se obtienen clústeres cada vez más pequeños y con mayor densidad.

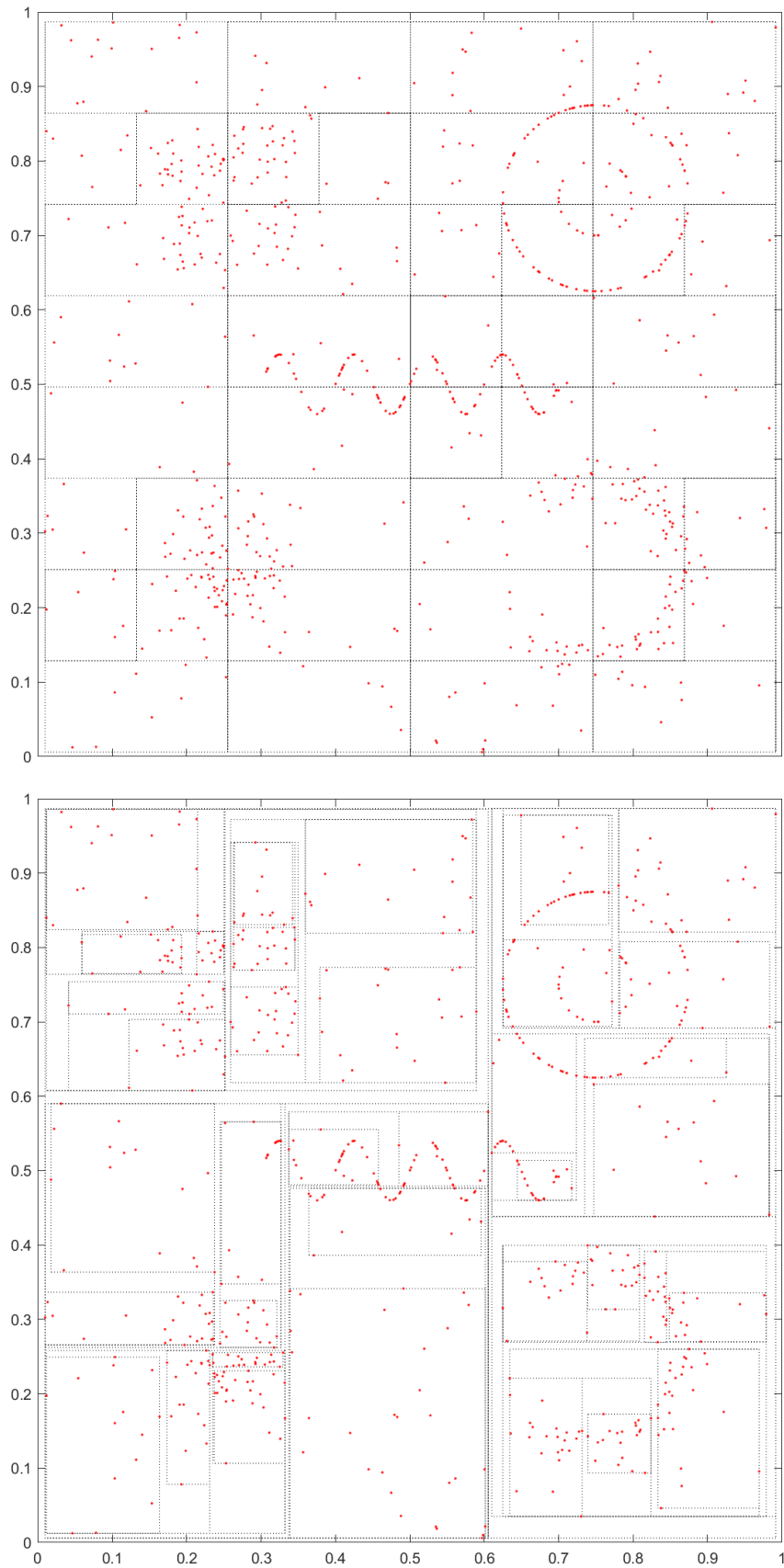


Fig. A.4: Visualización de las celdas en las que se dividen los árboles KD (arriba) y R (abajo). Se muestran superpuestos en rojo los puntos de la base de datos. El número máximo de puntos en una hoja es 30. El KD-tree obtenido tiene 83 nodos y el R-tree 73.

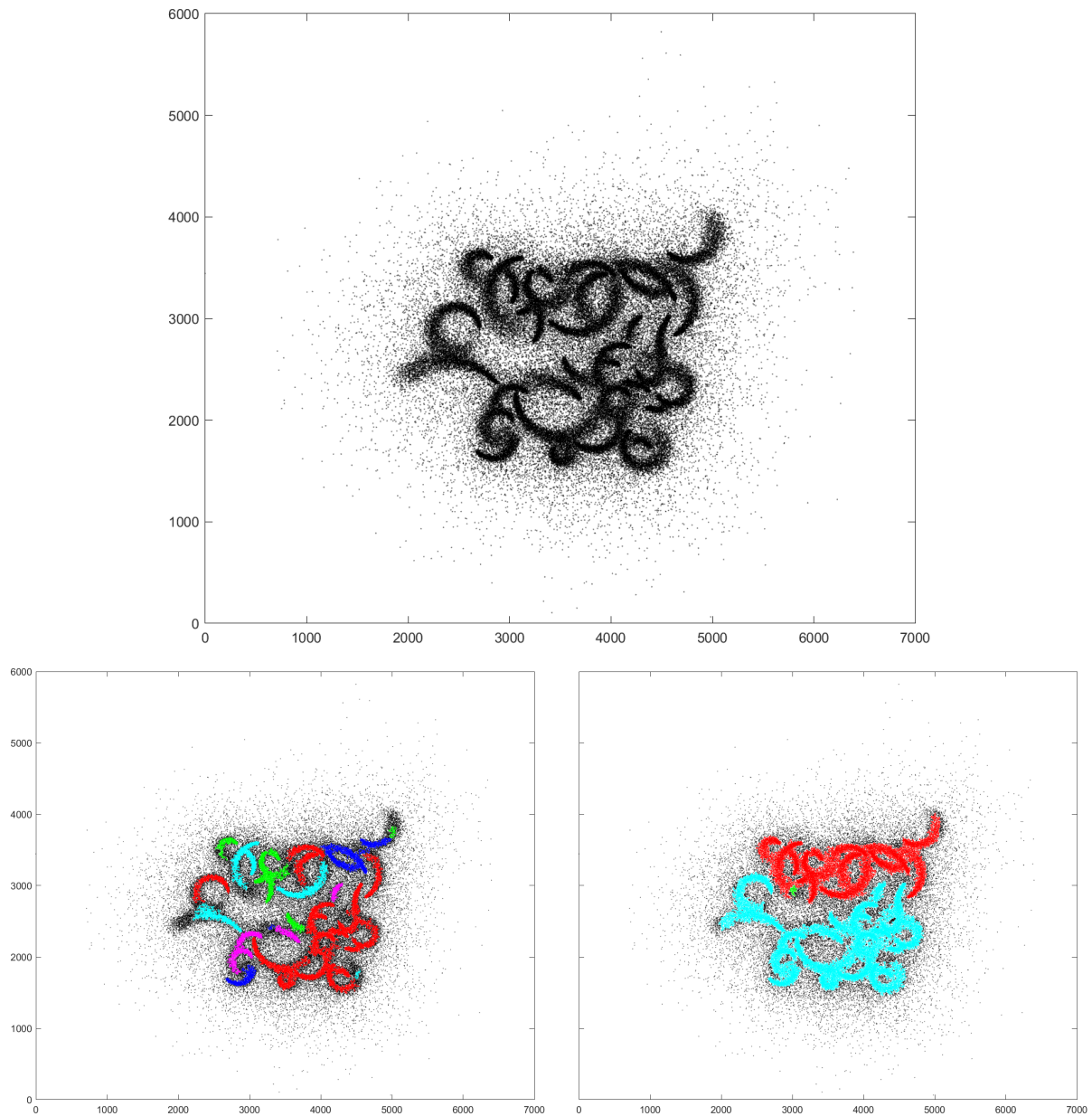


Fig. A.5: Clusters obtenidos con DBSCAN (izquierda) y DTFoF (derecha) para la base de datos worms (arriba). En los dos casos se ha usado una misma distancia máxima de enlace $b = 10$ y un número mínimo de puntos para formar un cluster $nMin = 100$. Para DBSCAN se ha tomado $minVecinos = 5$. DBSCAN es capaz de diferenciar algunos de los clusters de forma eficaz. Por el contrario, al haber tanta densidad de puntos, DTFoF tiende a agrupar los clusters. En total DBSCAN identifica 18 clústeres, mientras que DTFoF solamente 3.

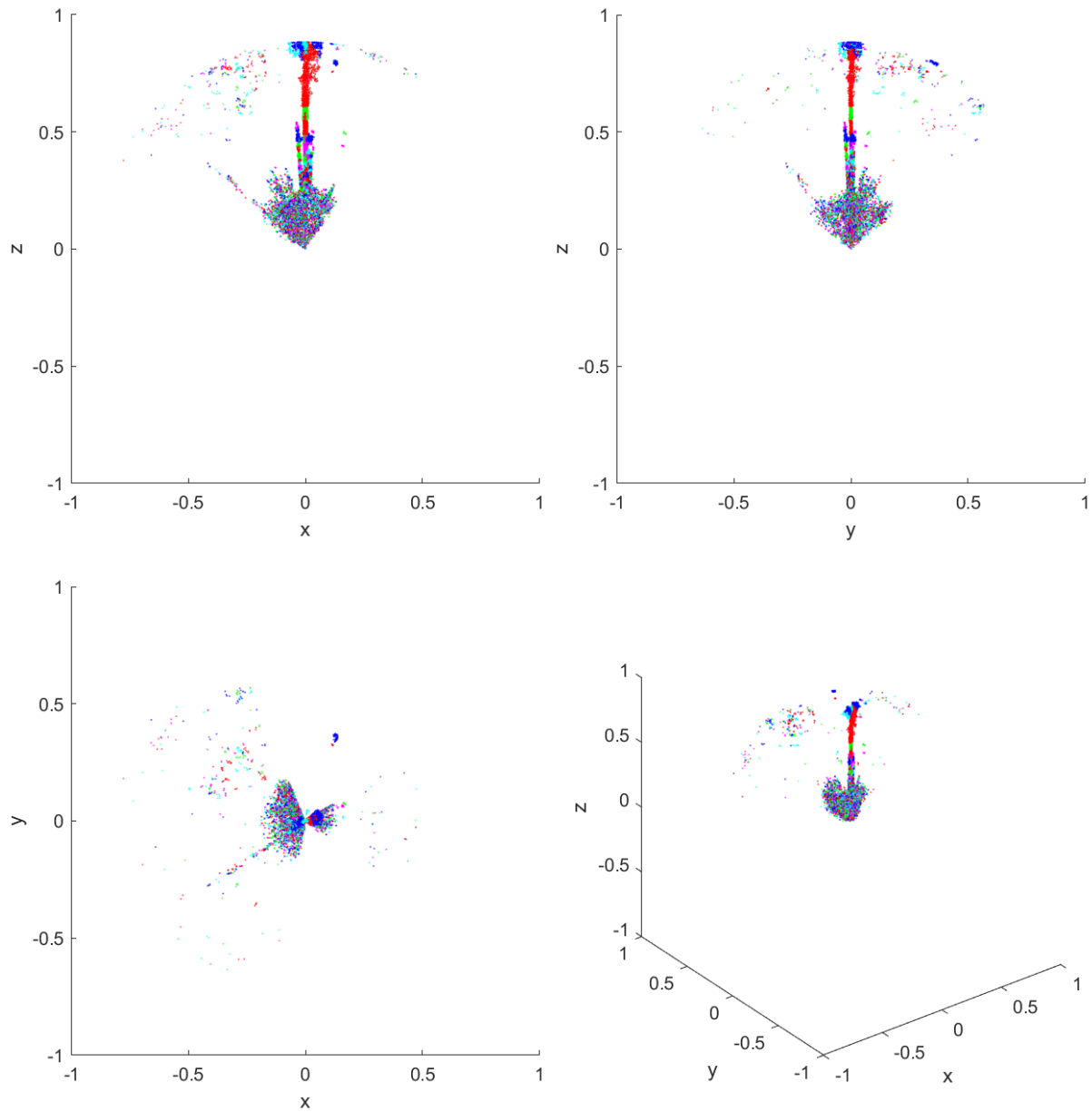


Fig. A.6: Clusters de galaxias tipo COMA obtenidos con DBSCAN en el sistema de coordenadas J2000. Se muestran solamente las galaxias que forman parte de un clúster. Todas las distancias están dadas en 10^4 MPc.

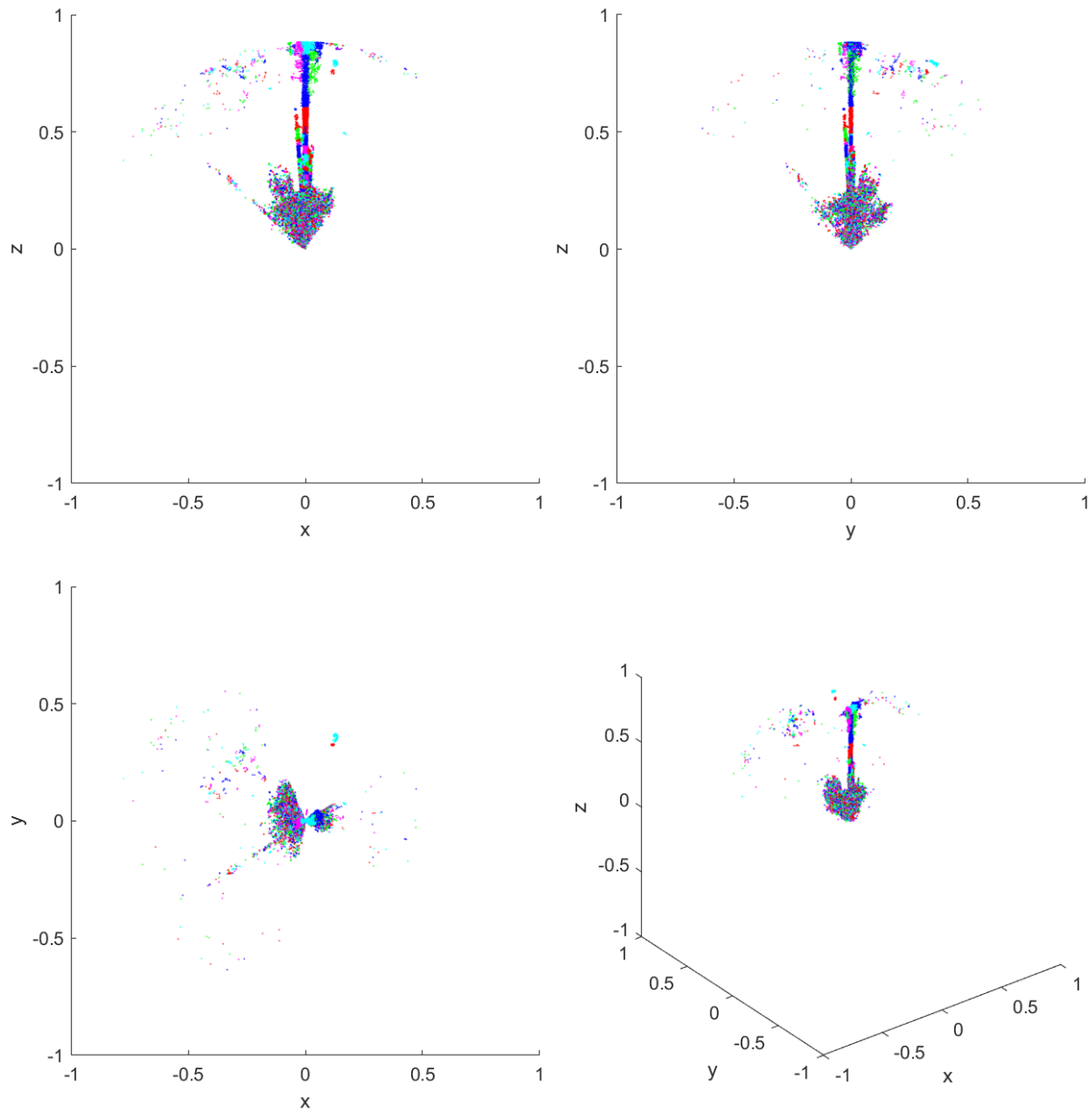


Fig. A.7: Clusters de galaxias tipo VIRGO obtenidos con DBSCAN en el sistema de coordenadas J2000. Se muestran solamente las galaxias que forman parte de un clúster. Todas las distancias están dadas en 10^4 MPc.

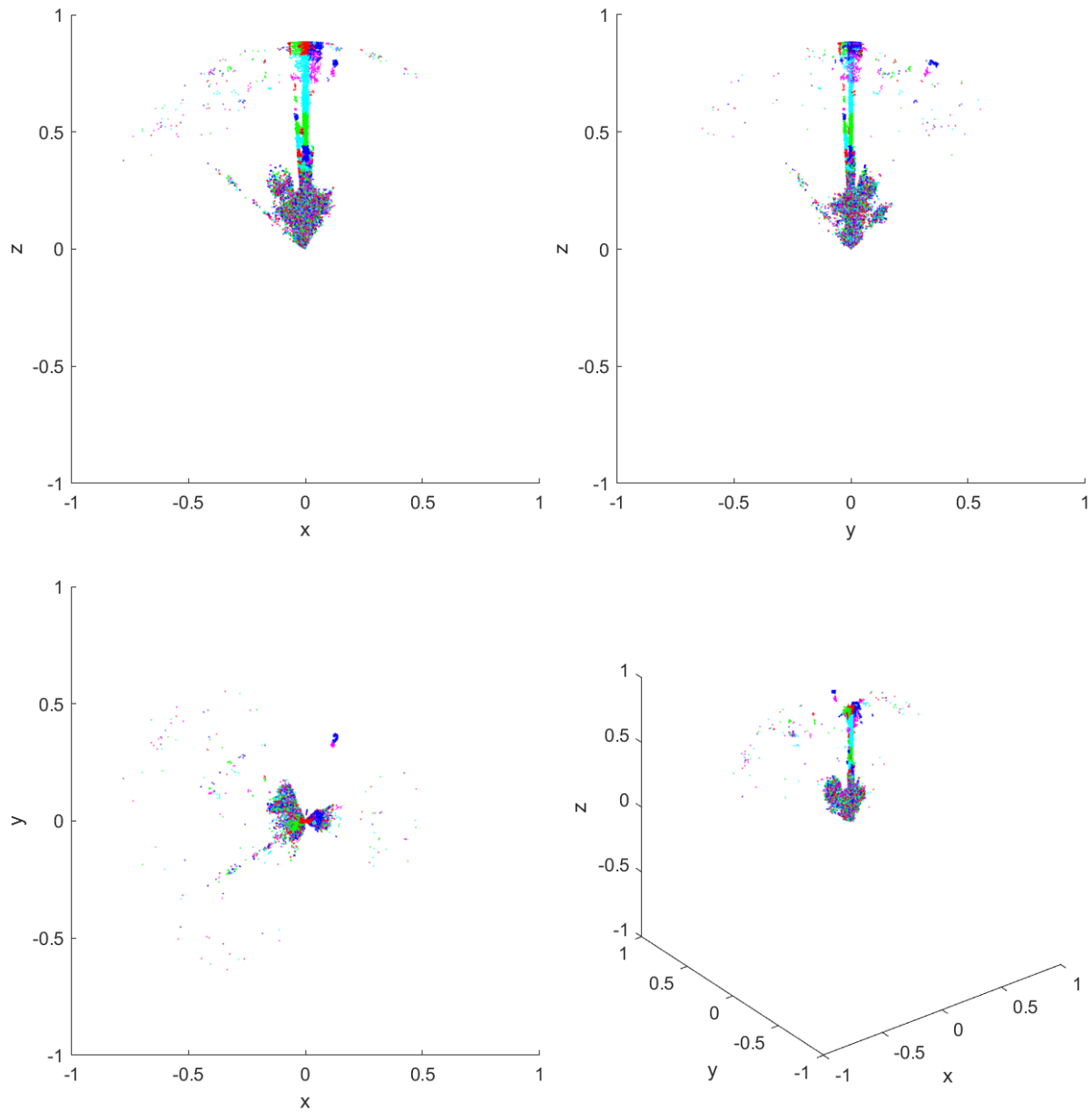


Fig. A.8: Clusters de galaxias tipo FORNAX obtenidos con DBSCAN en el sistema de coordenadas J2000. Se muestran solamente las galaxias que forman parte de un clúster. Todas las distancias están dadas en 10^4 Mpc.