



UNIVERSIDAD DE CONCEPCIÓN
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

Simulador de Fibras Cerebrales para validar los algoritmos de clustering de fibras y parcelación de la corteza basada en conectividad estructural

Por: Elida Rosa Poo Pérez

Tesis presentada a la Facultad de Ingeniería de la Universidad de Concepción
para optar al grado académico de Doctor en Ciencias de la Ingeniería con
Mención en Ingeniería Eléctrica

Noviembre 2024
Concepción, Chile

Tutora: Dra. Pamela Guevara
Co-Tutora: Dra. Cecilia Hernández

© 2024, Elida Rosa Poo Pérez

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento

AGRADECIMIENTOS

Quisiera agradecer a mi familia por todo el apoyo y comprensión durante este período y a las profesoras Pamela Guevara y Cecilia Hernández por su guía y orientación.

Esta tesis fue apoyada por:

- Agencia Nacional de Investigación y Desarrollo (ANID), Doctorado Nacional/21210353
- FONDECYT Regular 1190701
- FONDECYT Regular 1221665
- ANILLO ACT210053
- BASAL Centers FB0008 (AC3E)
- BASAL Centers FB210017 (CENIA)

Resumen

Las imágenes de difusión, combinadas con algoritmos de tractografía, permiten reconstruir los principales fascículos de materia blanca de manera *in vivo*. Diversos métodos han sido desarrollados para procesar los datos generados por la tractografía, con el fin de identificar las principales estructuras y conexiones cerebrales, junto con sus funciones. Entre estos métodos se encuentran el *clustering* de fibras y la parcelación cortical que son importantes en la neurociencia computacional y han sido la base de varios estudios de materia blanca. Sin embargo, estos métodos carecen de un *ground truth*. Una alternativa para evaluar la eficacia de estos algoritmos es el uso de conjuntos de datos de fibras simulados. No obstante, la simulación de las fibras cerebrales representa un desafío debido a su forma irregular y compleja.

El objetivo de esta tesis es implementar un simulador de fascículos cerebrales para validar los algoritmos de *clustering* de fibras y parcelación cortical basada en conectividad estructural. Se proponen dos estrategias de simulación: una basada en curvas exponenciales y otra en *splines*. La primera estrategia utiliza centroides y los extremos de las fibras se modifican con funciones exponenciales. La segunda usa el centroide, un modelo tubular para modelar el fascículo y *splines* para la representación de las fibras que lo componen. Los algoritmos fueron validados mediante la simulación de fascículos de un atlas de Materia Blanca Profunda. Los fascículos simulados se compararon con los del atlas, y ambos algoritmos mostraron distancias inter-fascículo bajas y altos porcentajes de intersección entre los fascículos originales y los simulados. Sin embargo, el algoritmo basado en *splines* superó al de las curvas exponenciales, obteniendo valores más altos en las medidas comparativas.

Para probar la utilidad del simulador se simularon datos de cerebro completo para evaluar el rendimiento de los algoritmos de *clustering*: QuickBundles y Fast Fiber Clustering. Los resultados indican que QuickBundles tiende a dividir menos y Fast Fiber Clustering tiende a fusionar menos. La agrupación correcta de los *clusters* por parte de ambos algoritmos disminuye, cuando se aumenta el número de fascículos. Además, los dos algoritmos de *clustering* muestran un comportamiento robusto frente a la permutación de los datos de entrada.

El simulador basado en curvas *splines* fue extendido para incluir formas más

complejas en los extremos, como las parcelas corticales. Este modelo mantiene la forma tubular en las regiones intermedias y central, mientras que las externas las modela según los vértices de las parcelas. Usando esta extensión del simulador de fascículos y una parcelación basada en distancia geodésica se generó una base de datos simulada de parcelaciones y sus conexiones. Esta base de datos se utilizó para evaluar el comportamiento de un algoritmo de parcelación basado en conectividad estructural. Finalmente, se realizó una comparación entre las parcelas generadas por el algoritmo de parcelación basado en conectividad y las obtenidas por el algoritmo basado en distancia geodésica utilizando el coeficiente de DICE y el Índice de Rand Ajustado, lo que resultó en aproximadamente 120 parcelas similares por hemisferio.

El simulador de fascículos cerebrales presentado en este trabajo representa una herramienta innovadora ya que no existe ninguna similar en el estado del arte. En el contexto actual, es el primer simulador de fascículos de fibras cerebrales capaz de evaluar algoritmos de análisis de tractografía cerebral utilizando datos realistas. Además, los códigos del simulador y la base de datos simulada se ponen a disposición de la comunidad científica, facilitando su acceso y uso para futuras investigaciones.

Keywords –Imágenes de difusión, Tractografía, Materia Blanca, Fascículos, Parcelación de la Corteza, *Clustering* de fibras

Índice general

AGRADECIMIENTOS	I
Resumen	II
Índice de tablas	VIII
Índice de figuras	XI
Acrónimos	XIX
1. Introducción	1
1.1. Introducción General	1
2. Revisión Bibliográfica	6
2.1. Anatomía del Cerebro	6
2.2. Estudios de la Materia Blanca <i>in-vivo</i>	9
2.2.1. Resonancia Magnética	9
2.2.2. Imágenes de Resonancia Magnética por Difusión	10
2.2.3. Tractografía cerebral	11
2.3. Métodos de Análisis de Tractografía cerebral	12
2.3.1. Clustering de fibras	12
2.3.1.1. Quickbundles	13
2.3.1.2. FFClust	14
2.3.2. Validación de los algoritmos de clustering de Fibras	15
2.3.3. Parcelación cortical	17
2.3.3.1. Validación de los Algoritmos de Parcelación Cortical	20
2.4. Modelación de los Fascículos de Materia Blanca	20
2.5. Discusión	24

3. Fundamentos de la Investigación	26
3.1. Hipótesis	26
3.2. Objetivo General	26
3.3. Objetivos Específicos	27
3.4. Alcances y Limitaciones	28
4. Materiales	29
4.1. Base de Datos	29
4.2. Atlas de Fibras de la Materia Blanca Profunda	30
5. Simulador de Fascículos usando curvas exponenciales	32
5.1. Métodos	32
5.1.1. Fascículos simulados usando curvas exponenciales basado en un DWM atlas	33
5.1.1.1. Identificación de sub-fascículos utilizando el algoritmo Kmeans mejorado	34
5.1.1.2. Generación de fascículos cilíndricos	35
5.1.1.3. Adición de las curvas exponenciales	36
5.1.2. Fascículos simulados en todo el cerebro para la validación del algoritmo de clustering QB	39
5.2. Resultados	41
5.2.1. Validación del simulador usando un DWM atlas	41
5.2.2. Fascículos simulados en todo el cerebro para la validación del algoritmo de clustering QB	42
5.3. Discusión	43
6. Simulador de fascículos usando curvas spline	46
6.1. Métodos	46
6.1.1. Simulador de fascículos usando curvas spline	47
6.1.1.1. Paso A: Inicialización de parámetros	48
6.1.1.2. Paso B: Construcción del modelo tubular	50
6.1.1.3. Paso C: Generación de las curvas spline	51
6.1.2. Validación del simulador de fascículos mediante un DWM atlas	52
6.1.3. Fascículos simulados en todo el cerebro para la validación de los algoritmos de clustering	54

6.2. Resultados	59
6.2.1. Validación del simulador usando un DWM atlas	59
6.2.2. Validación de algoritmos de clustering de fibras usando el simulador	63
6.2.3. Tiempo de ejecución y complejidad computacional del simulador	73
6.3. Discusión	75
7. Base de datos simulados para la validación de un algoritmo de parcelación estructural	78
7.1. Métodos	78
7.1.1. Parcelación aleatoria basada en distancia geodésica	80
7.1.2. Simulación de las conexiones entre parcelas	81
7.1.2.1. Preprocesamiento de los datos de entrada	82
7.1.2.2. Intersección de la tractografía con el mallado cortical	83
7.1.2.3. Obtención de los centroides que conectan un par de parcelas	83
7.1.2.4. Simulador de fascículos	85
7.1.3. Validación de un algoritmo de parcelación basado en conectividad estructural	88
7.2. Resultados	92
7.2.1. Generación de la base de datos de parcelas simuladas	92
7.2.2. Generación de la base de datos de conexiones simuladas	92
7.2.3. Validación del algoritmo de parcelación cortical usando la base de datos simulada	94
7.3. Discusión	100
8. Discusión y Conclusiones	101
8.1. Trabajo Futuro	103
8.2. Publicaciones	104
8.2.1. Conferencias	104
8.2.2. Revistas WOS	105
Referencias	106
A. Material Suplementario	115

A.1. Simulador de fascículos usando curvas splines	115
A.1.1. Modelo preliminar utilizando regiones de sección transversal elípticas	115
A.1.2. Simulación de fascículos de un atlas de materia blanca profunda	119
A.1.3. Simulación de conjuntos de datos de cerebro completo . . .	122
A.1.4. Cálculo de parámetros de FFClust	126
A.1.5. Métricas para la evaluación de los algoritmos de clustering Quickbundles, FFClust y QBX.	127

Índice de tablas

5.2.1. Distancia media en mm, desviación estándar y porcentaje de intersección entre los fascículos simulados y los fascículos del atlas.	41
5.2.2. Resultados de rendimiento del algoritmo QuickBundles (QB) con dos umbrales de distancia diferentes (10 y 15 mm) para los cuatro conjuntos de datos de fascículos simulados.	43
6.2.1. Resultados de métricas entre los fascículos simulados y los fascículos del atlas para el simulador usando curvas spline y el simulador basado en curvas exponenciales [72]: Simulador anterior I (simulación con cinco centroides), Simulador anterior II (simulación con un centroide), Distancia entre fascículos \pm Desviación estándar (mm) ($ID \pm STD$) y Porcentaje de intersección entre fascículos (%) (IP).	60
6.2.2. Caracterización del <i>ground truth</i> : El número de centroides utilizados para la simulación, la distancia mínima y media entre centroides, el número mínimo y máximo de fibras en los fascículos, el número total de fibras en el <i>ground truth</i> y el número de fascículos con cruzamientos.	63
6.2.3. Valores de las métricas para evaluar el rendimiento del algoritmo QB para los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se resaltan las métricas para el umbral con el mejor rendimiento.	66
6.2.4. Valores de las métricas para evaluar el rendimiento del algoritmo FFclust para los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se resaltan las métricas para el umbral con el mejor rendimiento.	67

6.2.5.Rendimiento de los algoritmos QB y FFClust en la recuperación de fascículos cruzados para diferentes umbrales (10, 12, 15 y 20 mm) y tres conjuntos de datos <i>ground truth</i> (100, 500 y 1000 fascículos): Porcentaje de <i>clusters</i> recuperados por el algoritmo (% de <i>clusters</i> cruzados recuperados).	68
6.2.6.Rendimiento del algoritmo QB para cinco permutaciones aleatorias de los tres conjuntos de datos de tractografía simulada. La tabla muestra los valores de media y desviación estándar de todas las métricas. En azul, se destacan las métricas para el umbral con el mejor rendimiento.	73
6.2.7.Rendimiento del algoritmo FFClust para cinco permutaciones aleatorias del conjunto de datos de tractografía simulada. La tabla muestra los valores de media y desviación estándar de todas las métricas. En azul, se destacan las métricas para el umbral con el mejor rendimiento.	74
7.2.1.Resultados de la parcelación basada en conectividad estructural al variar los parámetros del algoritmo: umbral de representatividad de sujetos thr_{rp} , umbral de probabilidad thr_{dc} , umbral de intersección thr_{it} . Parcelas finales (PF) por hemisferio: izquierdo (LH) y derecho (RH). Adicionalmente, se indica la cantidad de parcelas similares (PS) por hemisferio (DICE >0.5) y el valor del Índice de Rand Ajustado. La mejor combinación de parámetros se destaca en azul.	96
A.1.1Relación de radios para las cinco regiones de sección transversal elíptica (R1-R5) y relación de radios promedio para los 28 fascículos del atlas de DWM analizados. También se muestran los resultados para el porcentaje de intersección (%) (IP) y la distancia entre fascículos \pm desviación estándar (mm) (ID \pm STD) para el modelo tubular en comparación con el modelo preliminar utilizando secciones transversales elípticas. En azul, se destacan los ocho fascículos que presentan mejoras.	118

A.1.2	Valores de las métricas True Positives (TP), False Negatives (FN), False Positives(FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV). Estas métricas se utilizaron para obtener las métricas principales: Precision, Recall y Geometric Accuracy (Acc), necesarias para evaluar el rendimiento del algoritmo QB en los tres conjuntos de datos de tractografía simulada. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.	127
A.1.3	Valores de las métricas True Positives (TP), False Negatives (FN), False Positives(FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV). Estas métricas se utilizaron para obtener las métricas principales: Precision, Recall y Geometric Accuracy (Acc), necesarias para evaluar el rendimiento del algoritmo FFClust en los tres conjuntos de datos de tractografía simulada. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.	127
A.1.4	Valores de las métricas para evaluar el rendimiento del algoritmo QBX en los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.	127
A.1.5	Valores de las métricas True Positives (TP), False Negatives (FN), False Positives(FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV). En azul se destacan las métricas correspondientes al umbral con el mejor rendimiento.	128
A.1.6	Variación de los valores de las métricas True Positives (TP), False Negatives (FN), False Positives (FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV) al permutar los datos de entrada del algoritmo FFClust y los tres conjuntos de datos. La tabla muestra la media y la desviación estándar de todas las métricas. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.	129

Índice de figuras

2.1.1. Anatomía Básica de las células cerebrales	6
2.1.2. Conexión entre las regiones de materia gris y materia blanca m . . .	7
2.1.4. Tipos de fibras presentes en la materia blanca: fibras comisurales, fibras de asociación y fibras de proyección. Imagen adaptada de [98].	9
2.2.1. Imágenes T1 y T2 en corte axial. Imagen adaptada de [56].	10
2.2.2. Tipos de difusión del agua. a) Isotrópica no restringida, b) Isotrópica restringida, c) Anisotrópica. La difusión del tejido celular es la que se muestra en C) Imagen adaptada de [66].	11
2.2.3. Proceso de calculo del algoritmo de tractografía. a) Modelos de difusión local para el seguimiento de las fibras, b) Tractografía de cerebro completo, c) Conjunto de fascículos que conectan varias regiones. Imagen adaptada de [45].	12
2.3.1. Métrica de distancia y resultado del algoritmo QB a) Representación de la métrica de distancia MDF en ambas direcciones. Las fibras se muestran con líneas continuas, mientras que los pares de puntos de las dos fibras, cuyas distancias contribuyen a la métrica general, se destacan con líneas discontinuas. b) Ejemplo de la aplicación de la distancia MDF a un fascículo (en morado) y el centroide calculado por el algoritmo QB (en amarillo). Imagen adaptada de [29].	14
2.3.2. La metodología general de FFClust consiste de 4 pasos, Paso 1: Construcción de los clusters de puntos, Paso 2: Generación de los clusters de fibras preliminares, Paso 3: Reasignación de los clusters pequeños, Paso 4: Fusión de los clusters de fibras candidatos. Figura adaptada de [96].	16
2.4.1. Evolución del método de simulación de fibras cerebrales como strands. Figura adaptada de [9].	22

2.4.2. Ejemplo de fascículos simulados. A) Vista de las 200 fibras seleccionadas como centroides. B) Fascículos generados a partir de la traslación de los centroides de A. C) Fascículos de fibras resultantes tras la aplicación del método de <i>clustering</i> , para el 10 % de ruido. Figura adaptada de [40].	23
4.2.1. Diferentes vistas de todos los fascículos del atlas. Figura tomada de [39].	30
5.1.1. Esquema general del algoritmo de simulación utilizando curvas exponenciales. Los pasos principales del método incluyen: 1) Identificación de sub-fascículos utilizando Kmeans mejorado, 2) Simulación de fascículos cilíndricos, 3) Adición de curvas exponenciales. Los principales parámetros de un sub-fascículo son: r_{e1} y r_{e2} , definidos como los radios de los primeros y segundos extremos, r_c como el radio central, y x_{d1} y x_{d2} como la posición de inicio de la dispersión para cada extremo.	34
5.1.2. Esquema general de la generación de conjuntos de datos de fascículos simulados para validar el algoritmo <i>QuickBundles</i> (QB). a) El procesamiento de la tractografía del sujeto incluye la reducción de dimensionalidad con el algoritmo FFClust, el filtrado por longitud de fibra y la selección de las fibras más distantes de un conjunto aleatorio de fibras. b) El simulador utilizando curvas exponenciales muestra el conjunto representativo de centroides, un fascículo simulado basado en un centroide del conjunto y todos los fascículos simulados. c) La aplicación de <i>QuickBundles</i> muestra los centroides y los <i>clusters</i> obtenidos al aplicar QB al conjunto de datos simulado.	39
5.2.1. Resultados finales de la simulación para los tres fascículos del atlas, utilizando cinco sub-fascículos y el mismo número de fibras que los fascículos originales: el fascículo uncinado, el tracto corticoespinal y el fascículo arqueado. a) Dos vistas del atlas de fascículos originales, b) Centroides obtenidos del Kmeans mejorado, c) Fascículos cilíndricos obtenidos mediante la traslación de los centroides, y d) Dos vistas del fascículo simulado utilizando curvas exponenciales	42

5.2.2. Fascículos simulados para los dos conjuntos de centroides. a) Conjunto de 100 centroides, b) Tres fascículos simulados para radios en los extremos (r_{e1} y r_{e2}) en el intervalo de 5-10 mm (arriba) y 10-15 mm (abajo) para el conjunto de 100 centroides, c) Conjuntos de datos de fascículos simulados para el conjunto de 100 centroides y los dos rangos de radios en los extremos. d), e) y f): Las mismas visualizaciones que en a), b) y c), respectivamente, pero para el conjunto de 500 centroides.	44
6.1.1. Esquema general del simulador de fascículos. Paso A: Parámetros iniciales: A.1) Representación de 5 regiones transversales. Paso B: Construcción del modelo cilíndrico: B.1) Definición del primer punto periférico del círculo (P_{ri}), B.2) Definición de los puntos periféricos del círculo. Paso C: Generación de las curvas <i>spline</i> : C.1) Generación de los puntos de control de la <i>spline</i> dentro del modelo cilíndrico, C.2) Simulación del fascículo, con fibras definidas por las <i>splines</i>	49
6.2.1. Ejemplos de simulación de fascículos basados en el atlas de fascículos DWM [39]. Los fascículos originales del atlas se muestran a la izquierda (azul), y sus correspondientes fascículos simulados a la derecha (rojo). A) El fascículo uncinado izquierdo, B) El longitudinal inferior izquierdo y C) El fronto-occipital inferior. IP: Porcentaje de intersección entre fascículos (%) ID: Distancia entre fascículos (mm).	61
6.2.2. Ejemplos de diferentes tipos de cruzamientos existentes entre dos fascículos en los conjuntos de datos del <i>ground truth</i> . El fascículo rojo y el azul representan dos fascículos diferentes del <i>ground truth</i> . La línea negra representa un marcador de distancia de 10 mm. . .	64
6.2.3. Resultados obtenidos del <i>clustering</i> de fibras de QB y FFClust para los conjuntos de datos de tractografía simulados con 100 fascículos (I), 500 fascículos (II) y 1000 fascículos (III): A) Fascículos simulados originales. <i>Clusters</i> obtenidos por QB y FFClust para diferentes umbrales: B) 10 mm, C) 12 mm, D) 15 mm, E) 20 mm.	65

6.2.4. Un ejemplo del buen rendimiento de los algoritmos de <i>clustering</i> QB y FFClust al recuperar un <i>cluster</i> del <i>ground truth</i> . La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.	69
6.2.5. Un ejemplo del mal rendimiento de los algoritmos de agrupamiento QB y FFClust al recuperar un <i>cluster</i> del <i>ground truth</i> . La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.	70
6.2.6. Un ejemplo del rendimiento de los algoritmos de <i>clustering</i> QB y FFClust al recuperar un <i>cluster</i> del <i>ground truth</i> , donde FFClust tiene un mejor desempeño que QB. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.	71
6.2.7. Un ejemplo del rendimiento de los algoritmos de <i>clustering</i> QB y FFClust al recuperar un <i>cluster</i> del <i>ground truth</i> , donde QB tiene un mejor desempeño que FFClust. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En la parte media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.	72
6.2.8. Tiempo de ejecución para la simulación de un fascículo que contiene un número creciente de fibras, entre 100 y 1000 fibras, utilizando diferentes configuraciones de radios. Se usan cuatro configuraciones de radios, con radios más altos para las regiones transversales externas (r_1 y r_5), un radio más bajo para la región central (r_3) y radios intermedios para las regiones r_2 y r_4	75

7.1.1. Esquema general de la secuencia de algoritmos aplicados en este capítulo para la generación de una base de datos simulada, utilizada en la validación del algoritmo de parcelación cortical. A: Generación de la base de datos simulada de parcelaciones y sus conexiones, B: Aplicación y validación de un algoritmo de parcelación cortical basado en conectividad estructural. Los cuadrados representan los algoritmos aplicados. Los polígonos representan los datos de entrada de cada algoritmo (polígonos en rojo representan salida de un algoritmo y entrada de otro)	79
7.1.2. Diagrama del algoritmo de parcelación cortical basado en distancia geodésica. Representación en grafo del mallado de un sujeto: V representa los nodos del grafo y E las aristas. Para la inicialización de los centroides, se utiliza el algoritmo K-means++ para ubicar las semillas (rojo), y posteriormente, se emplea la distancia geodésica (magenta) para definir las parcelas finales.	80
7.1.3. Esquema general para la simulación de las conexiones. A) Datos de entrada, B) Preprocesamiento de los datos, C) Obtención de los centroides, D) Simulador de fascículos.	82
7.1.4. Ejemplos de centroides que conectan un par de parcelas corticales. A: fibra corta que conecta dos parcelas cercanas, B: fibra larga que conecta dos parcelas lejanas.	87
7.1.5. Diagrama general del proceso de validación de un algoritmo de parcelación cortical basado en conectividad estructural. A-C) Algoritmo de parcelación cortical. D) Comparación entre las parcelaciones simuladas y las generadas por el algoritmo basado en conectividad estructural.	88
7.2.1. Parcelación GeoSP visualizada en tres sujetos de la base de datos ARCHI. A la izquierda se encuentra la vista posterior, en el centro la vista dorsal y a la derecha la vista lateral.	93
7.2.2. Ejemplo de un fascículo simulado a partir del simulador. A) muestra el fascículo simulado. B) Muestra el fascículo simulado (verde) y las parcelas que conecta (azul).	93
7.2.3. Conexiones simuladas de tres sujetos de la base de datos. Colores aleatorios representan los diferentes fascículos que componen las conexiones simuladas.	94

- A.1.2 Ejemplo del modelo tubular propuesto (A) junto con un modelo preliminar utilizando secciones transversales elípticas (B) para el fascículo uncinado izquierdo. El centroide del fascículo se muestra con una línea gruesa negra. Las regiones de sección transversal de los extremos aparecen en magenta (RR=0.9) y púrpura (RR=0.9), la región de sección transversal central en marrón (RR=0.7), y las regiones de sección transversal intermedias están en cian (RR=0.8) y verde (RR=0.8). Los puntos centrales de las regiones de sección transversal están en azul, mientras que los puntos periféricos que definen los 8 sectores por región están en rojo. RR: Relación de radios elípticos. 117
- A.1.3 Fascículos simulados del atlas de fascículos DWM (Guevara et al., 2012) para el hemisferio izquierdo (primera columna) y el hemisferio derecho (segunda columna). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo. A) Arqueado anterior, B) Arqueado, C) Arqueado posterior, D) Fibras del cingulum, E) Tracto corticoespinal, F) Fornix, G) Fascículo fronto-occipital inferior. 119
- A.1.4 Fascículos simulados del atlas de fascículos DWM (Guevara et al., 2012) para el hemisferio izquierdo (primera columna) y el hemisferio derecho (segunda columna). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo. A) Fascículo longitudinal inferior, B) Radiaciones talámicas frontales, C) Radiaciones talámicas motoras superiores, D) Radiaciones talámicas occipitales, E) Radiaciones talámicas parietales superiores, F) Radiaciones talámicas temporales, G) Fascículo uncinado. . . . 120

A.1.5	Ejemplos de simulación de fascículos basados en el atlas de fascículos DWM (Guevara et al., 2012). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo, tanto para el modelo tubular como para el modelo preliminar utilizando secciones transversales elípticas. A) El fascículo uncinado izquierdo, B) El longitudinal inferior izquierdo, y C) El fronto-occipital inferior. IP: Porcentaje de intersección entre fascículos (%) ID: Distancia entre fascículos (mm).	121
A.1.6	Centroides utilizados para la simulación de los tres conjuntos de datos de referencia de 100 (A), 500 (B) y 1000 (C) fascículos. . . .	122
A.1.7	Diferentes vistas del conjunto de datos de referencia de 100 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda, E) Vista lateral derecha.	123
A.1.8	Diferentes vistas del conjunto de datos de referencia de 500 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda y E) Vista lateral derecha.	124
A.1.9	Diferentes vistas del conjunto de datos de referencia de 1000 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda y E) Vista lateral derecha.	125
A.1.10	Curvas del método del codo utilizadas para calcular el número de <i>clusters</i> en cada punto de fibra (0, 3, 10, 17, 20) para el MiniBatch Kmeans, que corresponde al primer paso del algoritmo FFClust. Los puntos seleccionados para los diferentes conjuntos de datos de referencia son: para 100 fascículos $Kp_{end} = 35$, $Kp_{inter} = 25$, $Kp_{cent} = 15$, 500 fascículos: $Kp_{end} = 50$, $Kp_{inter} = 35$, $Kp_{cent} = 25$, y 1000 fascículos: $Kp_{end} = 60$, $Kp_{inter} = 40$, $Kp_{cent} = 25$	126
A.1.11	Resultados del <i>clustering</i> de fibras obtenidos con los algoritmos QB y QBX para conjuntos de datos de tractografía simulada con 100 fascículos (I), 500 fascículos (II) y 1000 fascículos (III): A) Fascículos simulados originales. <i>Clusters</i> obtenidos por QB y QBX para diferentes umbrales: B) 10 mm, C) 12 mm, D) 15 mm, E) 20 mm	128

Acrónimos

Acc Exactitud Geométrica (Geometric Accuracy)

ARI Índice de Rand Ajustado (Adjusted Rand Index)

dMRI Imágenes de Resonancia Magnética por difusión (*Diffusion Magnetic Resonance*)

dODF Función de distribución de orientación de la difusión (*Diffusion Orientation Distribution Function*)

DTI Imágenes del Tensor de Difusión (*Diffusion Tensor Imaging*)

DWM Materia Blanca Profunda (*Deep White Matter*)

FFClust Fast Fiber Clustering

FN Falsos Negativos (False Negatives)

fODF Función de Distribución de Orientación de la Fibra (*Fiber Orientation Distribution Function*)

FP Falsos Positivos (False Positives)

GeoSP Parcelación basada en distancia geodésica (Geodesic Parcellation)

GM Materia Gris (*Gray Matter*)

HARDI Imágenes de Difusión de Alta Resolución Angular (*High Angular Resolution Diffusion Imaging*)

MMR Proporción máxima de coincidencia (Maximum Matching Ratio)

MRI Imágenes de Resonancia Magnética (*Magnetic Resonance Imaging*)

OMA Concordancia optimizada (Optimized Matched Agreement)

OS Puntaje de Traslape (Overlap Score)

PPV Sensibilidad de los clusters (Valor Predictivo Positivo)

QB Quickbundles

Sn Sensibilidad de los clusters (Clustering Wise Sensitivity)

SWM Materia Blanca Superficial (*Superficial White Matter*)

TP Verdaderos Positivos (True Positives)

WM Materia Blanca (*White Matter*)

Capítulo 1

Introducción

1.1. Introducción General

Durante décadas, los neurocientíficos han dedicado sus esfuerzos a modelar el cerebro y su actividad como una red altamente conectada, basándose en el concepto clave de esta área: el conectoma. Este término se refiere al mapeo detallado de las conexiones neuronales en el cerebro, proporcionando una representación integral de cómo las distintas regiones interactúan entre sí [88]. El estudio del conectoma no solo busca describir estas conexiones con precisión, sino también establecer una base de referencia que permita identificar anomalías de conectividad asociadas a diversos trastornos cerebrales. Este conocimiento es fundamental para avanzar en la comprensión de las patologías neurológicas y mejorar los diagnósticos y tratamientos clínicos.

El desarrollo de herramientas sofisticadas de Imágenes de Resonancia Magnética *MRI* como las Imágenes de Difusión de Alta Resolución Angular *HARDI*, han permitido obtener datos de mayor calidad y resolución, lo que resulta en una descripción más certera de la estructura cerebral a nivel macroscópico [16]. La resonancia magnética por difusión (dMRI) es una técnica no invasiva que mide el

movimiento aleatorio de las moléculas de agua en los tejidos [48]. La tractografía basada en dMRI permite reconstruir los principales trayectos de las fibras de la Materia Blanca (WM) del cerebro, brindando un enorme potencial para el estudio de la conectividad estructural cerebral a lo largo de la vida y en diversas condiciones [15, 58, 103]. Los datos de tractografía son complejos debido a la gran cantidad de fibras, la alta proporción de fibras con ruido, y la morfología intrincada de las conexiones cerebrales, lo que presenta desafíos significativos para su análisis [55, 101].

Existen diferentes métodos de análisis de tractografía como los métodos de *clustering* de fibras y de parcelación de la corteza. Estos son una parte importante en la descripción del conectoma. Los algoritmos de *clustering* de fibras tienen un papel importante por su capacidad para identificar automáticamente fibras similares y eliminar las que contienen ruido [69, 40, 29, 78, 44, 96, 50]. Estos métodos analizan un conjunto de fibras cerebrales y las agrupan en *clusters* que contienen fibras con formas y posiciones espaciales similares. Aunque la mayoría de los métodos exploratorios de agrupamiento no incorporan directamente información anatómica, suelen utilizarse como un paso dentro de un *pipeline* de procesamiento más amplio con un objetivo específico, como la creación de un atlas de fascículos de WM, la segmentación de fascículos de WM, el filtrado de fibras de WM, o la generación de parcelación basada en la difusión.

Los algoritmos de parcelación de la corteza cerebral permiten subdividir la corteza cerebral en un conjunto de regiones que representen la conectividad anatómica y funcional del cerebro. Estas parcelaciones permiten realizar estudios más detallados de la conectividad y la actividad cerebral, ya que dividen la corteza en áreas con características similares en cuanto a estructura, función o conectividad [60]. Son fundamentales para la creación de modelos de redes cerebrales [34][27][92], el estudio de enfermedades neurológicas [6, 86] y con esto el avance en la comprensión

de la organización funcional del cerebro.

Sin embargo, a pesar de su utilidad, estos algoritmos son difíciles de evaluar, comparar y mejorar debido a la ausencia de un *ground truth*. Diferentes algoritmos de *clustering* de fibras producen diferentes grupos de fibras a partir de un conjunto de datos de tractografía. Además, el mismo algoritmo genera diferentes agrupaciones dependiendo de los parámetros de entrada. De manera similar, los algoritmos de parcelación cortical también pueden producir resultados variados según los parámetros y enfoques utilizados. Por lo tanto, es fundamental desarrollar estrategias de evaluación robustas que permitan a los investigadores confirmar que los resultados de sus algoritmos son realmente sólidos y comparables con otros algoritmos a pesar de usar metodologías diferentes.

La literatura existente ofrece herramientas limitadas para evaluar y comparar diferentes resultados de agrupamiento de fibras y parcelación cortical en conjuntos de datos de tractografía cerebral completa. Respecto a los algoritmos de *clustering* se han utilizado múltiples métricas como por ejemplo el *Optimized Matched Agreement* (OMA), *Bundle Adjacency* (BA) [29], *Davies–Bouldin Index* (DB) [96, 33], *Precision*, *Recall*, *F-measure* [73] para medir la calidad de los *clusters* obtenidos. En la evaluación de los algoritmos de parcelación cortical se usan otras métricas como coeficiente de Dice [52, 82], el Índice de Rand Ajustado [2] y el coeficiente de silhouette [12]. En la mayoría de los casos, la metodología empleada para la evaluación varía según el grupo de investigación, y no existe un consenso general sobre cuáles son las técnicas óptimas para evaluar el desempeño de los algoritmos de *clustering* y parcelación cortical.

La obtención de datos simulados de fibras sería una alternativa eficaz para evaluar y caracterizar el desempeño de los algoritmos de *clustering* y parcelación. Sin embargo la simulación de las fibras cerebrales constituye un gran desafío, debido a que su forma es irregular y compleja, lo que dificulta lograr una representación

realista de las mismas.

En el estado del arte existen gran variedad de simuladores de fibras cerebrales, pero la mayoría fueron implementados para validar los algoritmos de tractografía y los modelos de difusión locales. Esto deriva en descripciones simples, que solo simulan la forma genérica de las posibles configuraciones de las fibras, como por ejemplo, las fibras que se cruzan de diferentes formas [75, 67]. Otros simuladores han estado dirigidos a análisis de variabilidad de fascículos para estudios poblacionales, o de cambios de la forma de los fascículos con la edad [32]. Estas limitantes en el estado del arte, sumadas a la importancia que representan los algoritmos de *clustering* y parcelación de la corteza cerebral para el estudio del conectoma conforman una de las principales motivaciones para el desarrollo de esta tesis.

Esta investigación se enfoca en el desarrollo de un simulador de fascículos de fibras basado en parámetros configurables por el usuario, que permita crear fascículos de apariencia realista de manera rápida y eficiente. Este simulador proporcionará una herramienta útil para la validación de algoritmos de *clustering* de fibras y de parcelación cortical. La implementación del simulador emplea curvas matemáticas para representar las fibras cerebrales. En una etapa preliminar del simulador, se utilizaron curvas exponenciales, que permiten una representación sencilla de fibras con poca curvaturas. En una etapa más avanzada, se incorporaron *splines* para representar fibras con curvaturas más complejas. El simulador se ha utilizado para evaluar el rendimiento de dos algoritmos de *clustering* existentes en el estado del arte Quickbundles y FFClust. Posteriormente el simulador basado en curvas *splines* mejorado para representar formas más complejas se utilizó para generar una base de datos simulada, con el fin de evaluar el desempeño de un algoritmo de parcelación cortical basado en conectividad estructural. No se encontraron evidencias en el estado del arte de una herramienta similar para la validación de algoritmos de *clustering* de fibras y parcelación cortical, lo que subraya la

importancia y novedad de esta investigación.

Capítulo 2

Revisión Bibliográfica

2.1. Anatomía del Cerebro

El cerebro humano es un sistema altamente complejo y organizado, es la estructura principal del sistema nervioso central y regula toda la actividad humana. Consta de varios cientos de miles de millones de neuronas como unidad funcional básica, que se comunican mediante múltiples interconexiones [89]. Todas las neuronas constan de cuatro partes distintas: cuerpo celular, dendritas, axones y terminales de axones como puede verse en la Figura 2.1.1.

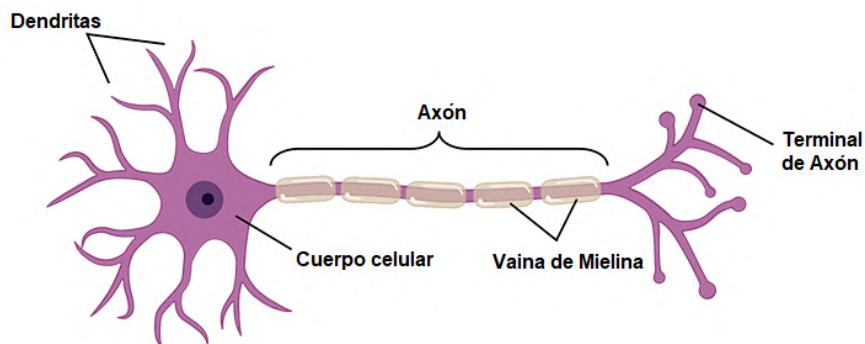


Figura 2.1.1: Anatomía Básica de las células cerebrales¹.

Los cuerpos de las células neuronales residen en la Materia Gris (*GM*) y los axones forman la materia blanca. Los axones a menudo se denominan fibras nerviosas y los conjuntos que forman estos axones se denominan tractos de fibras o fascículos. Las regiones de materia gris se encuentran conectadas por fibras de materia blanca que residen en el interior como se muestra en la Figura 2.1.2.

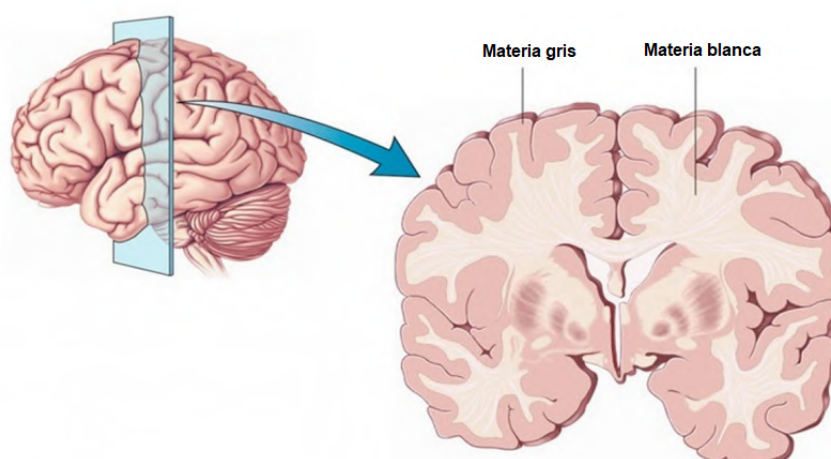


Figura 2.1.2: Conexión entre las regiones de materia gris y materia blanca².

La materia blanca forma parte de los lóbulos cerebrales y cerebelosos, además de los haces de fibras de conexión en el tronco encefálico y la médula espinal (ver Figura 2.1.3). Su función principal es conectar neuronas de diferentes áreas del cerebro mediante fibras nerviosas recubiertas de mielina, una sustancia lipídica que acelera la transmisión de impulsos eléctricos. Su característico color blanco brillante se debe a la alta concentración de lípidos en las vainas de mielina [13].

Los axones de materia blanca del cerebro están compuestos por tres tipos principales de fibras nerviosas: fibras comisurales, fibras de asociación y fibras de proyección, como pueden verse en la Figura 2.1.4.

¹source:<https://www.ninds.nih.gov/health-information/public-education/brain-basics/brain-basics-know-your-brain>

²source: <https://psicologiamymente.com/neurociencias/materia-gris-cerebro>

³source: <https://my.clevelandclinic.org/health/body/21598-brainstem>

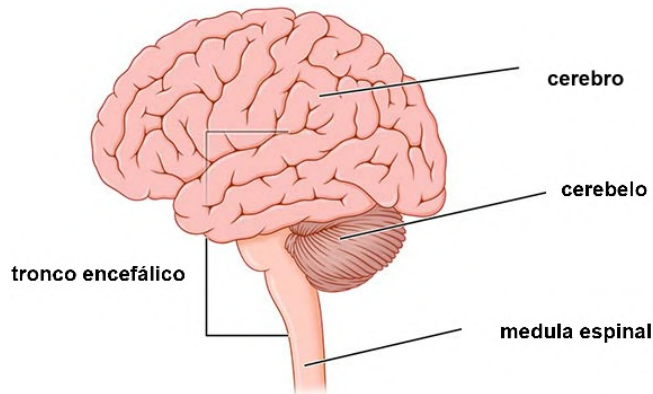


Figura 2.1.3: Conexión entre el cerebro, cerebelo, médula espinal y tronco encefálico³.

Las fibras comisurales conectan las regiones correspondientes de los hemisferios cerebrales izquierdo y derecho. Las fibras de asociación conectan regiones de la corteza dentro del mismo hemisferio. Hay dos tipos de fibras de asociación: fibras cortas y fibras largas. Las fibras de asociación cortas se pueden encontrar justo debajo de la corteza y en la superficie de la materia blanca. Estas fibras conectan las circunvoluciones cerebrales. Las fibras de asociación largas conectan los lóbulos cerebrales dentro de las regiones del cerebro. En el caso de las fibras de proyección, estas conectan la corteza cerebral con el tronco encefálico y la médula espinal. Estos tractos de fibras ayudan a transmitir señales motoras y sensoriales entre el sistema nervioso central y el sistema nervioso periférico (SNP) [23]. El SNP comprende todos los nervios y ganglios fuera del sistema nervioso central, conectando el cerebro y la médula espinal con el resto del cuerpo para facilitar la comunicación bidireccional[65].

La materia gris es un tejido donde se lleva a cabo el procesamiento de información, siendo responsable de la integración y transmisión de señales neuronales para generar respuestas cognitivas y motoras. En los hemisferios cerebrales, la materia gris se sitúa en la capa más externa, denominada corteza [63]. La composición

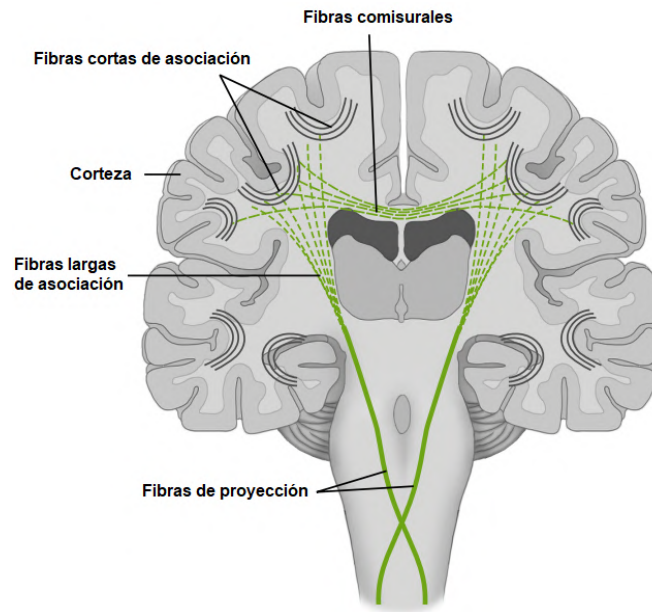


Figura 2.1.4: Tipos de fibras presentes en la materia blanca: fibras comisurales, fibras de asociación y fibras de proyección. Imagen adaptada de [98].

principal de la corteza cerebral es de materia gris y tiene una apariencia rugosa. Consta de protuberancias, también conocidas como *gyri* (circunvoluciones), y surcos profundos, conocidos como *sulci* (surcos), de entre 2 y 3 milímetros de grosor, representando el 40% de la masa cerebral. También se divide en dos hemisferios, el derecho y el izquierdo, que están separados por un gran surco. Los dos hemisferios están conectados a través del cuerpo calloso.

2.2. Estudios de la Materia Blanca *in-vivo*

2.2.1. Resonancia Magnética

La resonancia magnética (MRI) es una técnica no invasiva que utiliza campos magnéticos y señales de radiofrecuencia para obtener imágenes detalladas de estructuras anatómicas, como el cerebro. Su principio funcional se basa en la interacción entre los protones en los tejidos y el campo magnético aplicado, lo que

genera señales que son procesadas para formar imágenes [36]. Dos parámetros claves en esta técnica son T1 (spin-lattice relaxation time) y T2 (spin-spin relaxation time), que describen el tiempo que tardan los tejidos en regresar al equilibrio tras un pulso de radiofrecuencia. T1 está relacionado con la interacción de los protones con su entorno, mientras que T2 refleja la interacción entre los protones. Estos parámetros, junto con la densidad de protones (PD), determinan el contraste y la visibilidad de las estructuras en las imágenes obtenidas (ver Figura 2.2.1). En este estudio las imágenes T1 se utilizan para obtener imágenes anatómicas del cerebro mientras que las T2 están asociadas a imágenes de difusión que permiten la reconstrucción de la Materia Blanca cerebral.

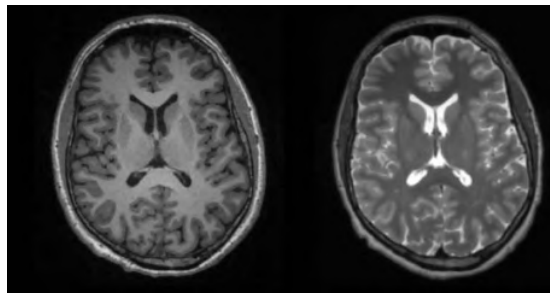


Figura 2.2.1: Imágenes T1 y T2 en corte axial. Imagen adaptada de [56].

2.2.2. Imágenes de Resonancia Magnética por Difusión

Las imágenes de resonancia magnética por difusión (*dMRI*) permiten el mapeo del proceso de difusión en tejidos biológicos de manera no invasiva, a través de la medición de la difusión restringida de las moléculas de agua [71].

La difusión de las moléculas de agua en espacio libre tiene un movimiento llamado Browniano, que consiste en un movimiento aleatorio en todas las direcciones[48]. En cambio, para la Materia Blanca (WM) el área restringida por las membranas del axón es mucho más acotada debido a las vainas de mielina, dificultando el movimiento en todas las direcciones y fortaleciendo el movimiento a lo largo del

axón. Esto se traduce en una difusión altamente anisotrópica como se muestra en la Figura 2.2.2, revelando la orientación local de las fibras en cada vóxel (unidad tridimensional mínima de un volumen de imagen) y posibilitando la reconstrucción de las fibras cerebrales en forma de curvas tridimensionales [3].

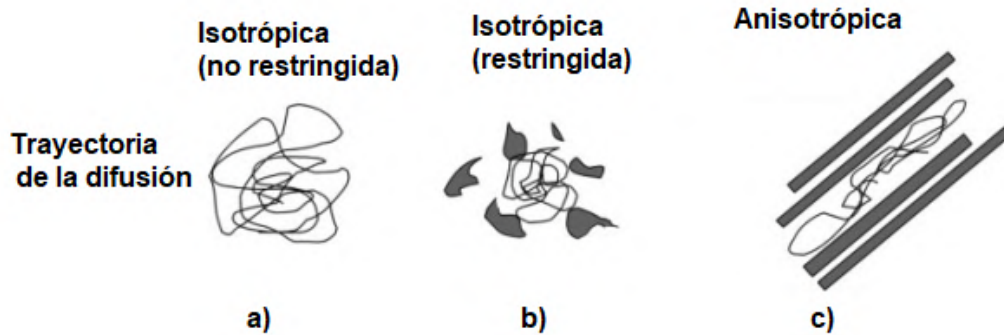


Figura 2.2.2: Tipos de difusión del agua. a) Isotrópica no restringida, b) Isotrópica restringida, c) Anisotrópica. La difusión del tejido celular es la que se muestra en C) Imagen adaptada de [66].

2.2.3. Tractografía cerebral

A la información direccional proporcionada por la dMRI se le pueden aplicar distintos modelos locales de difusión para así describir tridimensionalmente el desplazamiento molecular del agua durante la secuencia de difusión [17, 91] (ver la Figura 2.2.3.a parte superior).

Usando los modelos de difusión se calcula la tractografía, que no es más que una técnica de reconstrucción 3D [64] que permite reconstruir las principales vías de los axones dentro de la *WM* cerebral, como un conjunto de curvas conocidas como *streamlines* o fibras cerebrales [3], ver la Figura 2.2.3.b. Las fibras cerebrales pueden agruparse en estructuras denominadas fascículos según un patrón espacial definido por la arquitectura anatómica y funcional [83] como se muestra en la Figura 2.2.3.c.

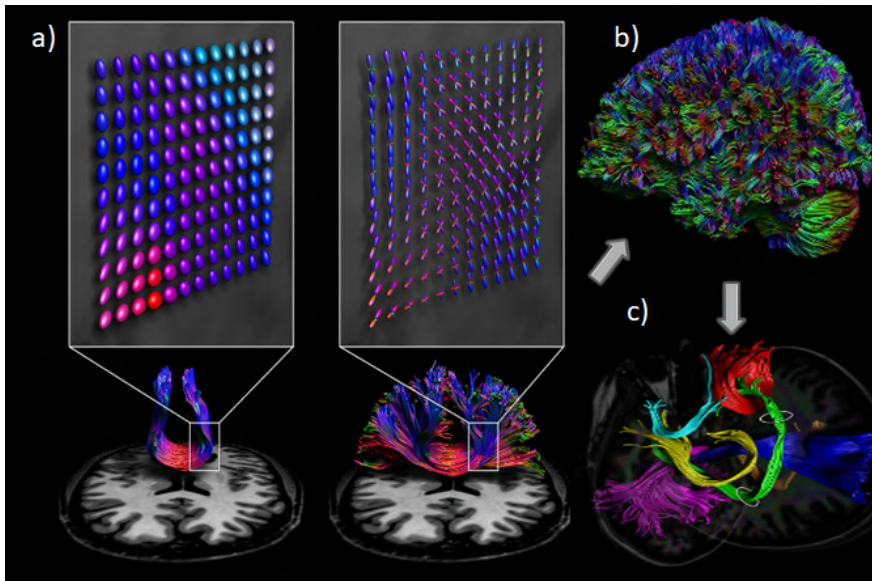


Figura 2.2.3: Proceso de cálculo del algoritmo de tractografía. a) Modelos de difusión local para el seguimiento de las fibras, b) Tractografía de cerebro completo, c) Conjunto de fascículos que conectan varias regiones. Imagen adaptada de [45].

2.3. Métodos de Análisis de Tractografía cerebral

Al estimar y reconstruir la trayectoria de las fibras por medio de un algoritmo de tractografía, se obtiene como resultado un conjunto de datos no etiquetados. Este conjunto de datos contiene fibras que pertenecen a fascículos conocidos, y también puede contener fibras ruidosas y artefactos, procedentes de las limitaciones intrínsecas e incertidumbres de la *dMRI* [96]. Utilizando estos datos es posible aplicar métodos como el *clustering* de fibras para establecer cómo se agrupan las fibras en fascículos y métodos de parcelación de la corteza para el estudio de patrones de conectividad entre sujetos.

2.3.1. Clustering de fibras

La aplicación de métodos de *clustering* ha ayudado a desarrollar estudios de los fascículos de Materia Blanca Profunda (*DWM*) debido a su capacidad para identificar automáticamente fibras similares y descartar fibras ruidosas [69, 78, 44,

96, 50]. En particular han contribuido a la construcción de atlas como los propuestos por [70] y [39]. Más recientemente, se han llevado a cabo investigaciones acerca de la Materia Blanca Superficial (*SWM*) o fibras de asociación corta, partiendo de la reproducibilidad de estos fascículos de fibras y la construcción de dos atlas de fascículos de la *SWM*, el presentado en Guevara et al. (2017) y el propuesto en Román et al. (2017).

En general, el *clustering* es un algoritmo de aprendizaje no supervisado dedicado al agrupamiento de elementos en función de un criterio de similitud. En el caso de las fibras cerebrales, se busca agrupar fibras en *clusters* o centroides que ayuden a representar la estructura de la conectividad cerebral [29, 40]. Como medida de similitud en la *WM* generalmente se utiliza una distancia entre fibras que considera su posición y forma.

Aunque la mayoría de los métodos de *clustering* exploratorio no incorporan directamente información anatómica, normalmente se utilizan como un paso en un proceso de procesamiento más extenso con un objetivo específico, por ejemplo, crear un atlas de fascículos de Materia Blanca (*WM*), realizar una segmentación de fascículos de *WM*, filtrar fibras de *WM* o generar una parcelación basada en difusión.

2.3.1.1. Quickbundles

Uno de los métodos de *clustering* más populares en el estado del arte es *QuickBundles* (*QB*) [29]. *QB* mide la similitud entre fibras utilizando la distancia mínima promedio directa e inversa (*MDF*, Minimum Average Direct Flip), como se ilustra en la Figura 2.3.1.a. Las principales ventajas de la distancia *MDF* son su rapidez en el cálculo y la capacidad para considerar la direccionalidad de las fibras. Para garantizar que las fibras sean comparables, *MDF* requiere que tengan el mismo número de puntos. Dependiendo de la aplicación, se pueden utilizar más

o menos puntos. En general, se realiza un remuestreo a 12 puntos, ya que ofrece un buen equilibrio entre la resolución de las fibras y la reducción del tamaño de memoria [29, 30].

El algoritmo comienza seleccionando la primera fibra en la tractografía y asignándola al primer *cluster*. Posteriormente, utiliza un umbral de distancia (θ_{QB}) para decidir si una nueva fibra debe asignarse al *cluster* más cercano o iniciar uno nuevo. QB ha sido empleado en diversas aplicaciones, tales como la segmentación de fascículos de materia blanca [31] y el filtrado de fibras [22]. Un ejemplo de la distancia MDF aplicada a un fascículo y el cálculo del centroide en QB se muestra en la Figura 2.3.1.b.

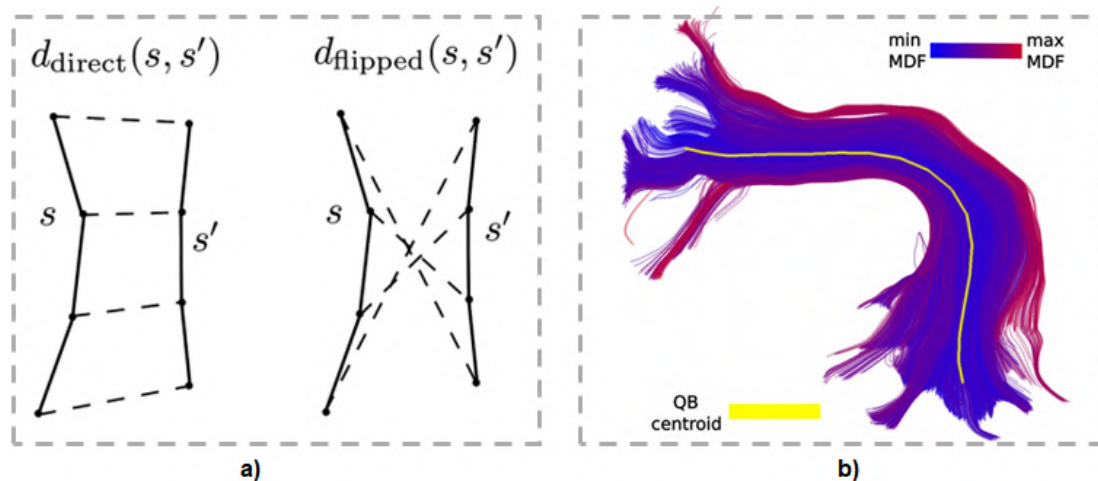


Figura 2.3.1: Métrica de distancia y resultado del algoritmo QB a) Representación de la métrica de distancia MDF en ambas direcciones. Las fibras se muestran con líneas continuas, mientras que los pares de puntos de las dos fibras, cuyas distancias contribuyen a la métrica general, se destacan con líneas discontinuas. b) Ejemplo de la aplicación de la distancia MDF a un fascículo (en morado) y el centroide calculado por el algoritmo QB (en amarillo). Imagen adaptada de [29].

2.3.1.2. FFClust

Otro algoritmo de agrupación de fibras es *Fast Fiber Clustering* (FFClust) [96], que fue diseñado principalmente para identificar *clusters* compactos en conjuntos

de datos masivos de tractografía con un tiempo de computación reducido. Este algoritmo consta de cuatro pasos. Primero reduce la dimensionalidad de los datos aplicando un MiniBatch Kmeans [85] sobre un subconjunto de puntos de fibras. En el segundo paso, las fibras cuyos puntos comparten los mismos *clusters* de puntos se agrupan en *clusters* de fibras preliminares.

En el tercer paso, los *clusters* pequeños se reasignan al *cluster* grande más cercano, teniendo en cuenta un umbral de distancia (dR_{max}). Si, al final del paso, sigue habiendo *clusters* formados por sólo una o dos fibras, éstos se consideran ruidosos y se descartan. En el último paso los centroides candidatos de *clusters* se fusionan si la distancia Euclidiana, directa o inversa, entre ellos está por debajo del umbral de distancia (dM_{max}), usando una representación de grafo no dirigido donde cada centroide es un vértice. En el grafo se identificaron los cliques máximos formados por centroides cercanos. Un clique es un subconjunto de vértices en el que cada par está directamente conectado por una arista. Un clique máximo, por su parte, es aquel que no permite agregar ningún vértice adicional sin perder esta propiedad [97]. Posteriormente, estos cliques se ordenan y fusionan si contienen al menos dos vértices. Una representación de los cuatro pasos se muestra en la Figura 2.3.2. Este algoritmo se ha utilizado para la creación de un atlas de fascículos de Materia Blanca Superficial[81], un método para la parcelación cortical basada en la difusión [61], y la eliminación de valores atípicos.

2.3.2. Validación de los algoritmos de clustering de Fibras

La validación de los algoritmos de *clustering* es una tarea difícil, ya que existen una gran variedad de métodos, cuyos algoritmos de agrupamiento están basados en diferentes características de las fibras; pero sobre todo porque no existe un *ground truth*. Debido a esto, los investigadores han desarrollado diferentes técnicas que permitan evaluar sus resultados.

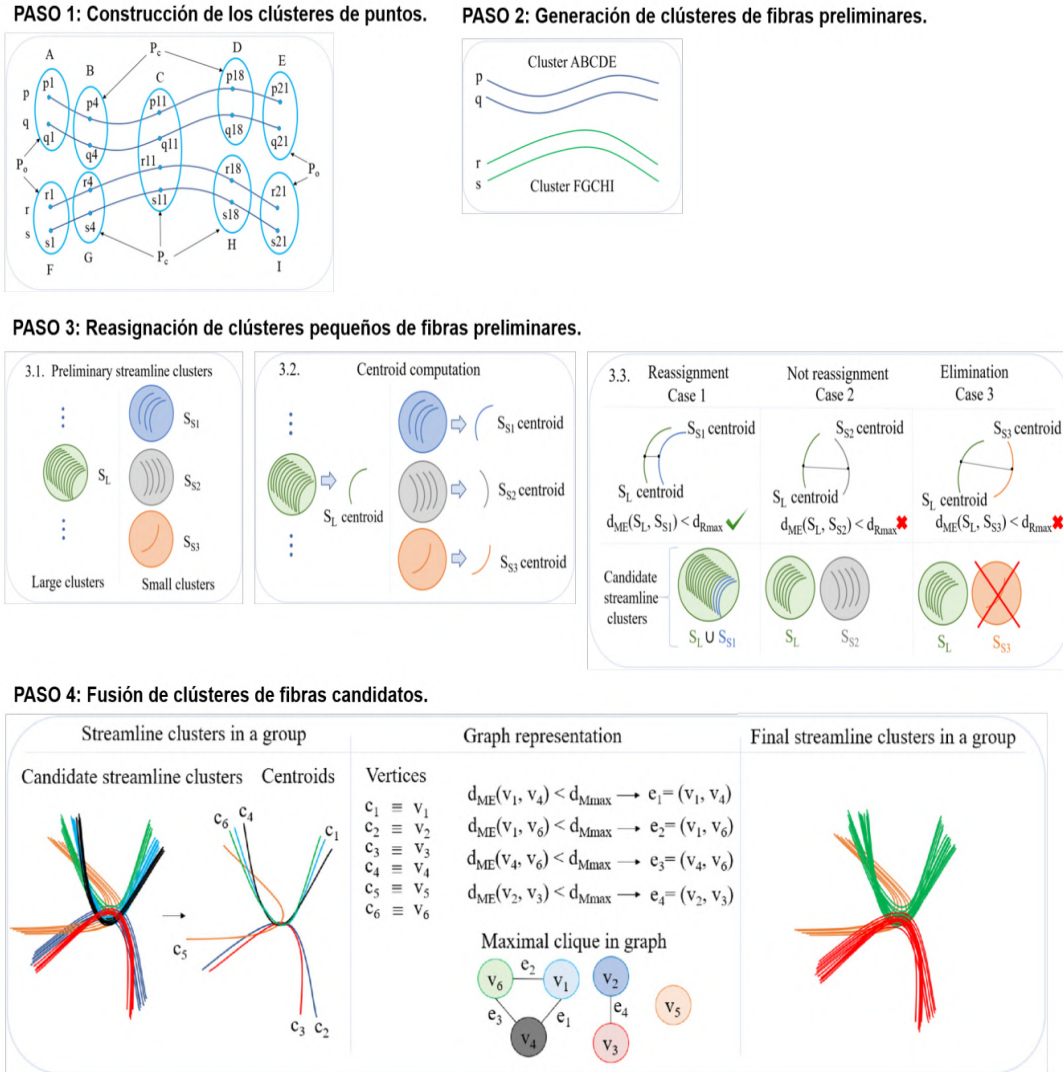


Figura 2.3.2: La metodología general de FFClust consiste de 4 pasos, Paso 1: Construcción de los clusters de puntos, Paso 2: Generación de los clusters de fibras preliminares, Paso 3: Reasignación de los clusters pequeños, Paso 4: Fusión de los clusters de fibras candidatos. Figura adaptada de [96].

Por ejemplo, los autores de QuickBundles utilizaron diferentes conjuntos de datos de tractografía de cerebro completo con condiciones iniciales y cuatro métricas para evaluar su algoritmo. Estas métricas son la *Optimized Matched Agreement (OMA)*, *Coverage*, *Overlap*, y *Bundle Adjacency (BA)*. OMA se utilizó para comparar los diferentes *clusters* que surgen cuando se aplican permutaciones aleatorias al conjunto de datos de fibras de entrada. Las métricas de *coverage* y *overlap* se

calcularon para evaluar en qué medida los centroides QB son una mejor reducción que un número equivalente de selecciones aleatorias de fibras. Por último, se empleó el BA para comparar los resultados de *cluster* entre sujetos. Los autores realizaron una simulación sencilla de fibras que generó tres conjuntos distintos de fibras a partir de combinaciones de funciones sinusoidales y helicoidales, lo que ilustró el comportamiento del algoritmo para umbrales bajos y altos.

Por otro lado, los autores del algoritmo FFClust [96] proporcionaron una evaluación del rendimiento con otros métodos del estado del arte [40, 29, 30]. La calidad de los *clusters* obtenidos por cada algoritmo se evaluó utilizando las distancias intracluster e intercluster y el índice Davies-Bouldin (DB) [100], definido como la similitud media entre cada *cluster* con su *cluster* más similar. Además, compararon los tiempos de ejecución de los algoritmos y determinaron que su implementación paralela es unas 8.6 veces más rápida que QB.

Otros han utilizado clasificación manual de las fibras realizada por expertos para comparar si los fascículos generados por su método concuerdan con estructuras reales [59]. En Guevara et al. (2011b) se utilizó un conjunto de datos de fibras simulados para evaluar la homogeneidad de los fascículos resultantes de un *clustering* jerárquico. También se han utilizado otros enfoques de métricas de evaluación como la precisión, el recall y el F1 score [44]. Aunque las métricas y comparaciones utilizadas han permitido una evaluación parcial del comportamiento de estos algoritmos, sería deseable disponer de una herramienta de simulación que permitiera la validación con datos que proporcionen un *ground truth* a escala de todo el cerebro.

2.3.3. Parcelación cortical

Un paso fundamental en el mapeo macroscópico del cerebro, así como en el análisis de sus conexiones estructurales y funcionales, es la subdivisión del cerebro en

regiones a gran escala. Este proceso, que organiza la corteza cerebral en áreas basadas en características específicas, se conoce como parcelación cortical [14]. Según el método empleado para dividir la corteza existen diferentes parcelaciones: basadas en la composición celular de la corteza (citoarquitectónica), en regiones específicas asociadas a distintas funciones cognitivas (funcional), dependientes de la conectividad axonal (estructural), y por último pueden integrar información de diversas metodologías de neuroimagen (multimodal) [60]. La parcelación estructural es la que se estará analizando durante el desarrollo de esta tesis.

En una parcelación estructural, cada parcela presenta un patrón de conectividad distinto con respecto a un conjunto predefinido de regiones de interés. La forma más habitual de generar una parcelación estructural es calcular primero una matriz de conectividad entre regiones y, a continuación, parcelar mediante alguna técnica de *clustering*. Las técnicas más populares para parcelar la matriz de conectividad estructural son umbralización, Kmeans, análisis de componentes principales, análisis de componentes independientes, *clustering* espectral y el *clustering* jerárquico [28].

Entre algunos de los trabajos destacados que utilizan esta técnica se encuentra Behrens et al. (2003), donde a partir de la tractografía, se calcula como conectan un conjunto de vóxeles semilla en el tálamo, con regiones corticales definidas anatómicamente. A continuación, asignan a cada vóxel semilla la etiqueta de la región cortical con la que está más conectada. Anwander et al. (2007) utiliza el *clustering* Kmeans sobre la matriz de conectividad del área de Broca, obteniendo una división en tres regiones, consistente con las divisiones citoarquitectónicas. Mientras que en Gallardo et al. (2018) parcelan el lóbulo frontal en 12 regiones mediante el análisis de componentes principales, y muestran su reproducibilidad entre sujetos.

Otros han propuesto métodos híbridos para crear parcelaciones de la superficie

cortical [52], partiendo de una parcelación gruesa basada en un atlas anatómico y utilizando información de conectividad estructural proporcionada por fascículos segmentados de una población de sujetos. El método toma como entrada la tractografía, los mallados etiquetados de los sujetos según las regiones del atlas anatómico Desikan-Killiany [18] (los vértices están etiquetados según la región a la cuál pertenecen en el atlas) y un atlas fusionado de fascículos de materia blanca superficial y profunda. Como resultados se genera un atlas promedio que subdivide las parcelas anatómicas del atlas Desikan-Killiany, identificando subparcelas estables basadas en la conectividad entre los sujetos.

Vergara et al. (2021) propone un algoritmo para la parcelación cortical a nivel multi-sujeto, basado en la conectividad estructural calculada a partir de *clusters* de fibras cerebrales representativos de una población de sujetos. El método inicia con un *clustering* intra-sujeto usando el algoritmo FFClust para obtener los *clusters* específicos de cada individuo. Posteriormente, se realiza un *clustering* inter-sujeto sobre estos *clusters* para identificar los representativos de la población, en función de un umbral de representatividad, thr_{rp} . La creación de la parcelación final se realiza en cuatro pasos. En el primer paso, se definen las regiones preliminares: cada *cluster* que intersecta la malla cortical en sus extremos crea dos parcelas preliminares, etiquetadas con el nombre del *cluster* y un sufijo *A* o *B* según el extremo correspondiente. En el segundo paso, para cada triángulo y su vecindad considerando todos los sujetos, se calcula la densidad de intersecciones con diferentes fascículos y la probabilidad de cada parcela. En el tercer paso, las parcelas con superposición significativa en áreas densas se fusionan según dos umbrales. Primero, un umbral de densidad thr_{dc} se utiliza para identificar el centro de cada parcela. Luego, si el grado de intersección entre parcelas supera un umbral de intersección thr_{it} , estas se agrupan en cliques máximos dentro de un grafo, asignando etiquetas de parcelas definitivas. Finalmente, en el cuarto paso, se aplica

una apertura morfológica (erosión y dilatación) para eliminar triángulos poco conectados y rellenar huecos, logrando así una parcelación cortical homogénea y continua.

2.3.3.1. Validación de los Algoritmos de Parcelación Cortical

Evaluar la calidad de los métodos de parcelación es una tarea compleja, ya que no existe un *ground truth* universal para la parcelación de la corteza cerebral. En la literatura, las técnicas de evaluación varían dependiendo de los aspectos específicos de la parcelación que se analicen [2]. Algunas de las métricas más utilizadas incluyen el coeficiente Dice [52, 82, 49], que mide la superposición entre dos parcelaciones, y el Índice de Rand Ajustado [49, 2], que evalúa la concordancia entre ellas. Otras métricas adicionales incluyen la homogeneidad [12], definida como la similitud de los vértices dentro de una misma parcela usando el coeficiente de correlación de Pearson, que mide la relación lineal entre dos variables. Mientras que el coeficiente Silhouette [12, 35] evalúa qué tan bien se ajustan los vértices a su parcela considerando tanto la cohesión interna de la parcela (diferencia promedio de distancia entre un vértice y otros de su misma parcela) como la separación respecto a otras parcelas.

2.4. Modelación de los Fascículos de Materia Blanca

A pesar de los avances en el conocimiento sobre la composición de la materia blanca, la forma geométrica de sus fibras sigue siendo menos estudiada, y su validación se ha realizado principalmente a través de estudios *post mortem* y la construcción de atlas anatómicos basados en la forma general y la ubicación de los fascículos. Sin embargo, la morfología de los fascículos aún ofrece oportunidades

de investigación inexploradas.

Varios estudios se han enfocado específicamente en la geometría de los fascículos de la materia blanca. En [Batchelor et al. \(2006\)](#), se propuso el uso de invariantes geométricos, como la curvatura y la torsión, para describir la forma de los fascículos a lo largo de su recorrido. Estas métricas permiten cuantificar la geometría de las fibras sin perder detalles esenciales sobre su configuración, como su forma intrínseca y las relaciones espaciales que mantienen entre sí a lo largo de su recorrido. De manera similar, en [Durrleman et al. \(2011\)](#), se presenta un método para representar las fibras como rutas integrales sobre un campo vectorial, lo que permite comprimir los datos sin perder su integridad geométrica, ajustando los parámetros según la complejidad de los datos.

Otro estudio desarrolló un *framework* más avanzado para ser utilizado en la validación del algoritmo de tractografía, pero no permiten sustituir a los datos *in vivo* [67]. Este *framework* es de código abierto y define de forma intuitiva los tractos de fibra; a partir del estudio de parámetros microestructurales de la WM para la representación utiliza el modelo Fibercup [75][74][24]. El modelo Fibercup es un fantoma físico compuesto por fibras acrílicas dispuestas en patrones específicos para replicar las complejas configuraciones de la materia blanca cerebral. Estas fibras simulan estructuras como el cruzamiento, donde las fibras se interceptan en forma de “X” sin fusionarse; el contacto o *kissing*, en el que las fibras se aproximan en paralelo sin cruzarse; la fusión, en la que varios fascículos se combinan y recorren una trayectoria común; y la división, en la que un fascículo se bifurca en diferentes direcciones [26]. Al estar hechas de acrílico, estas fibras poseen propiedades estructurales que permiten replicar con precisión la dirección y disposición de los fascículos cerebrales en un entorno controlado. Esto facilita a los investigadores la evaluación de algoritmos de tractografía en un contexto físico que emula las configuraciones del cerebro humano, mejorando así la validación de estos métodos

en condiciones más realistas [75].

Close et al. (2009) propuso una herramienta interesante que utiliza una colección de construcciones numéricas conocidos como strands (hilos). Este enfoque consiste en simular fascículos como strands gruesos que se representan mediante *splines* lineales en 3D con secciones transversales circulares constantes. La simulación comienza con la inicialización de estos strands a lo largo de segmentos de línea recta, conectando puntos generados aleatoriamente en la superficie de una esfera, hasta que toda la esfera queda cubierta. Se utiliza una función de costo para controlar el grado de superposición. Esta forma de representación se muestra en la Figura 2.4.1.

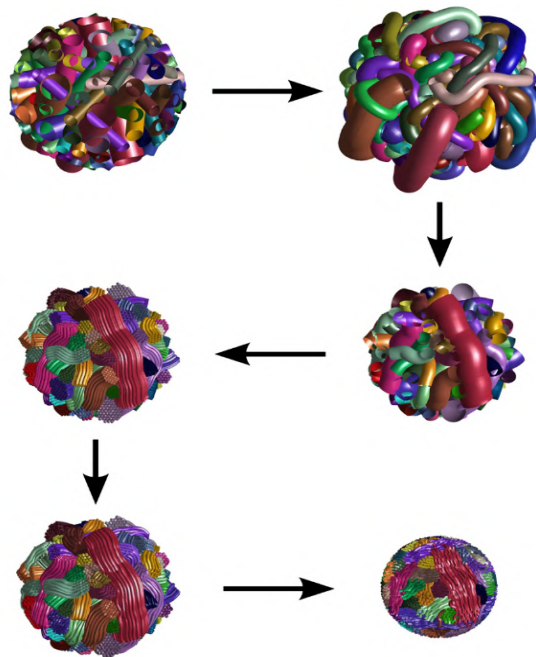


Figura 2.4.1: Evolución del método de simulación de fibras cerebrales como strands. Figura adaptada de [9].

Otros trabajos realizan el análisis de los fascículos de fibras utilizando datos de secciones transversales para representar la forma de los fascículos. Un ejemplo fue propuesto por Glozman et al. (2018), quienes desarrollaron un marco de trabajo para evaluar los cambios en la forma de los fascículos dependientes de la edad

entre distintos sujetos. Para lograrlo, utilizaron un modelo geométrico que divide el fascículo en secciones transversales y extrae un conjunto de puntos semilla, capturando de manera precisa su forma. Finalmente los puntos de las fibras se agrupan en *clusters* en función de su ubicación espacial y su geometría local mediante vecinos más cercanos. Con esta información se encuentran y se analizan los cambios en las estructuras.

Por otro lado, el trabajo de [Guevara et al. \(2011b\)](#) tiene un enfoque diferente, ya que propuso una simulación de fascículos de fibras para validar su propio algoritmo de *clustering* jerárquico de fibras. Los datos simulados se generaron a partir de un conjunto de centroides, definidos por fibras seleccionadas aleatoriamente del conjunto de datos de tractografía de cerebro completo de un sujeto. Los fascículos se formaron mediante una traslación aleatoria de sus centroides. El resultado fue un conjunto de fascículos cilíndricos distribuidos por todo el cerebro, con un número variable de fibras añadidas para simular ruido, como se muestra en la [Figura 2.4.2](#).



Figura 2.4.2: Ejemplo de fascículos simulados. A) Vista de las 200 fibras seleccionadas como centroides. B) Fascículos generados a partir de la traslación de los centroides de A. C) Fascículos de fibras resultantes tras la aplicación del método de *clustering*, para el 10% de ruido. Figura adaptada de [40].

2.5. Discusión

Las Imágenes de Resonancia Magnética por Difusión es una técnica no invasiva que mide el movimiento aleatorio de las moléculas de agua en los tejidos. La tractografía de fibras basada en dMRI permite reconstruir las principales vías de fibras de la materia blanca del cerebro, ofreciendo un enorme potencial para el estudio de la conectividad estructural del cerebro. No obstante, los datos derivados de la tractografía son intrínsecamente complejos debido al gran número de fibras involucradas, la presencia significativa de fibras ruidosas, y la morfología altamente compleja de las conexiones cerebrales, lo que plantea importantes desafíos analíticos.

Existen varios métodos para analizar los datos de tractografía, incluyendo algoritmos de agrupamiento de fibras, que han hecho una importante contribución debido a su capacidad de identificar automáticamente fibras similares y descartar fibras ruidosas. Otro método relevante para el análisis de tractografía son los algoritmos de parcelación cortical basados en la conectividad estructural, que resultan esenciales para comprender la función y organización cerebral, así como para investigar trastornos neurológicos.

A pesar de su relevancia, la evaluación, comparación y mejora de estos algoritmos sigue siendo un reto debido a la ausencia de un *ground truth* claramente definido. Una alternativa prometedora consiste en el uso de datos simulados de tractografía, que podrían proporcionar un *ground truth* robusto para validar dichos algoritmos. Sin embargo, la simulación de fibras cerebrales es una tarea compleja debido a la naturaleza irregular y no lineal de las trayectorias de las fibras, lo que hace difícil generar datos que reflejen con precisión la realidad anatómica. Actualmente, la mayoría de los marcos de simulación disponibles están diseñados principalmente para validar algoritmos de tractografía o modelos de difusión local, pero no ofrecen

soluciones completas para simular el rango completo de complejidades presentes en los datos reales de fibras cerebrales.

Capítulo 3

Fundamentos de la Investigación

3.1. Hipótesis

La implementación de un simulador de fascículos de fibras cerebrales realistas, al considerar parámetros de la forma del fascículo, basado en el uso de herramientas matemáticas como las curvas exponenciales o *splines* para la representación de las fibras, permitirá evaluar el comportamiento de los algoritmos de *clustering* y parcelación de la corteza.

3.2. Objetivo General

Implementar un simulador de fascículos de fibras para validar algoritmos de análisis de tractografía cerebral, tales como el *clustering* de fibras y la parcelación cortical. El simulador debe ser capaz de generar fascículos con apariencia realista a partir de parámetros configurables, que sean característicos de la forma geométrica de los fascículos como los radios y la dispersión de los extremos.

3.3. Objetivos Específicos

OBJ 1: Desarrollar un algoritmo que sea capaz de simular fascículos de fibras con un aspecto realista, a partir de la selección de centroides y de la configuración de parámetros fundamentales de la forma del fascículo, como los radios, la posición de la dispersión y el número de fibras. Se usarán para la simulación de fibras cerebrales dos estrategias de simulación, utilizando curvas exponenciales y curvas *splines*.

OBJ 2: Demostrar la aplicabilidad del simulador de fascículos de fibras a la evaluación y caracterización del desempeño de un algoritmo de *clustering* de fibras. Para ello se simularán conjuntos de datos de fibras con distintos parámetros, a los cuales se les podrá adicionar fibras ruidosas, y se aplicará el algoritmo de *clustering*. Para analizar el comportamiento de este algoritmo, se calculará el porcentaje de fascículos originales que fueron recuperados al 100 % y el promedio de división de los *clusters*.

OBJ 3: Crear una base de datos simulados de parcelaciones corticales que contenga las regiones corticales y sus conexiones para una población de sujetos. La simulación de las regiones se realizará a partir de una parcelación basada en distancia geodésica sobre el mallado de un cerebro, y las conexiones se obtendrán a partir del simulador de fibras.

OBJ 4: Demostrar la aplicabilidad de la base de datos simulados de parcelaciones para evaluar un algoritmo de parcelación cortical basado en conectividad estructural. El algoritmo se evaluará analizando el solapamiento entre las parcelas generadas por este y las simuladas, usando el coeficiente de DICE y el índice de Rand Ajustado como métricas comparativas.

3.4. Alcances y Limitaciones

En esta tesis, para la simulación de los fascículos, se utilizaron los datos de tractografía de la base de datos HARDI denominada ARCHI [84]. Esta base de datos contiene información de 79 sujetos, adquirida en un escáner de MRI Tim Trio 3T, e incluye imágenes estructurales T1. Además, la base de datos proporciona los mallados corticales correspondientes a cada sujeto, así como las matrices de transformación afín para convertir entre los espacios T1 (espacio anatómico del sujeto), T2 (espacio de difusión del sujeto) y Talairach. Para la validación de los algoritmos de *clustering* de fibras, se utilizó únicamente la tractografía de un sujeto, mientras que para la validación de los algoritmos de parcelación se emplearon los datos de tractografía y los mallados de 20 sujetos de esta base de datos.

Se utilizó para la evaluación del simulador el atlas de fibras largas propuesto en [41]. Es un atlas de fascículos multi-sujeto *HARDI* que incluye 36 fascículos de *DWM*.

Capítulo 4

Materiales

En esta sección se presenta la base de datos utilizada para la simulación de los fascículos y el atlas de fascículos de la Materia Blanca Profunda utilizado para la validación del simulador.

4.1. Base de Datos

Se dispone de una base de datos HARDI de 79 sujetos sanos, denominada ARCHI [84]. La base de datos fue adquirida en un escáner de MRI Tim Trio 3T con una antena de cabeza de 12 canales (Siemens, Erlangen). El protocolo de adquisición incluyó la adquisición de un conjunto de datos de imágenes T1 utilizando una secuencia MPRAGE (160 cortes, TE/TR = 2,98/2300 ms, TH = 1,10 mm, ángulo de deflexión FA = 9, TI = 900 ms; matriz = 256 × 240; tamaño de voxel = 1 × 1 × 1,1 mm; RBW = 240 Hz/píxel), un conjunto de datos HARDI SS-EPI de una sola shell con 60 direcciones DW optimizadas, y un mapa de campo B₀, b = 1,500 s/mm² (70 cortes, TE = 93 ms, FA = 90, TH = 1,7 mm, TR = 14,000 ms, matriz = 128 × 128, factor de Fourier parcial PF = 6/8, espaciado de eco ES = 0,75 ms, RBW = 1,502 Hz/píxel; GRAPPA = 2, tiempo total de escaneo

= 16 minutos y 46 segundos). Además, están disponibles los mallados corticales correspondientes a cada sujeto, así como las matrices de transformación afín, que permiten convertir entre los espacios T1, T2 y Talairach.

Los datos fueron procesados con el software BrainVISA/Connectomist-2.0 [20]. Se corrigieron los principales artefactos, como corrientes eddy, efectos de susceptibilidad, distorsiones geométricas y ruido, y se eliminaron los cortes defectuosos. Luego se calculó el modelo de difusión q-ball analítico [16] y la tractografía determinística de tipo *streamline*, sobre una máscara de propagación del cerebro completo, basada en T1 [38], usando una semilla por vóxel (a resolución T1) y un paso de 0.1mm. Esto generó un conjunto de datos de tractografía con un promedio de un millón de fibras para cada sujeto.

4.2. Atlas de Fibras de la Materia Blanca Profunda

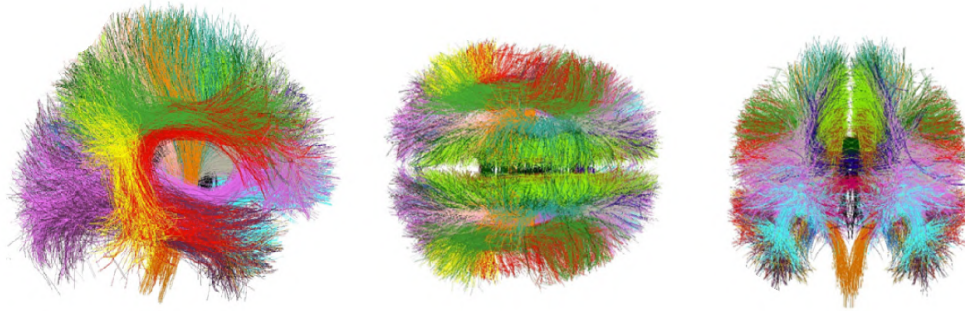


Figura 4.2.1: Diferentes vistas de todos los fascículos del atlas. Figura tomada de [39].

Para la evaluación del simulador se utilizó el atlas de fibras largas propuesto en [41]. Es un atlas de fascículos multi-sujeto *HARDI*, que representa la variabilidad de la forma y la posición de los fascículos entre sujetos. El atlas incluye 36 fascículos de *WM* profunda 18 por hemisferio (11,755 fibras). Estos fascículos fueron obtenidos utilizando la base de datos ARCHI. Finalmente el atlas está

compuesto por fascículos conocidos como el tracto cortico espinal, el fascículo uncinado, el arqueado, las fibras del cingulum, entre otros.

Capítulo 5

Simulador de Fascículos usando curvas exponenciales

En general, los fascículos de fibras cerebrales tienen una forma compleja, donde los extremos tienden a estar muy dispersos. Por lo tanto, es necesario desarrollar un método capaz de representarlos lo más realista posible. Este capítulo presenta un simulador de fascículos de fibras que usa un centroide y curvas exponenciales para la simulación de los fascículos. Este simulador es una primera aproximación de un modelo de simulador de fibras cerebrales.

5.1. Métodos

Esta sección presenta la metodología utilizada para generar un simulador de fascículos de fibras usando curvas exponenciales. Para simular un fascículo basado en un modelo, se subdividió el fascículo del atlas en sub-fascículos, aplicando un agrupamiento Kmeans. Luego, cada sub-fascículo es representado por un conjunto de curvas exponenciales con parámetros extraídos del sub-fascículo original. Este procedimiento permite obtener una simulación más realista.

Para validar el simulador, se utilizaron tres fascículos del atlas de fascículos de DWM [39] como modelo, los cuales tienen formas representativas diferentes: el fascículo uncinado, el tracto corticoespinal y el fascículo arqueado. Para la evaluación del simulador se calcularon las métricas distancia interfascículo y el porcentaje de intersección entre el fascículo del atlas y el fascículo simulado.

5.1.1. Fascículos simulados usando curvas exponenciales basado en un DWM atlas

Este algoritmo genera fascículos cilíndricos mediante la traslación de una o más fibras centroides N veces. Si se trabaja con varios centroides, estos se seleccionan utilizando un algoritmo K-means. En caso de utilizar un único centroide, este se calcula como el promedio de todas las fibras que componen el fascículo.

Posteriormente, a las fibras trasladadas se les adiciona curvas exponenciales para introducir dispersión en sus extremos. El efecto de la dispersión busca representar el comportamiento de las fibras al conectar la corteza luego de atravesar la materia blanca. Esta distribución de conexiones en los fascículos tienden a ser más compactas en las zonas internas de la corteza y más dispersas en las áreas periféricas. Por lo tanto, el fenómeno de la dispersión busca hacer más realista la simulación del fascículo.

Las curvas exponenciales van a depender de parámetros como los radios externos y centrales de los fascículos, lo que asegura una apertura configurable en ambos extremos. Para esta simulación con fines de validación, los valores de estos parámetros se obtuvieron a partir de los fascículos originales del atlas DWM.

No obstante, los parámetros del radio, la posición de dispersión y el número de sub-fascículos son ajustables por el usuario según las necesidades de su aplicación. Un esquema general del proceso se presenta en la Figura 5.1.1, y los pasos específicos

se describen a continuación.

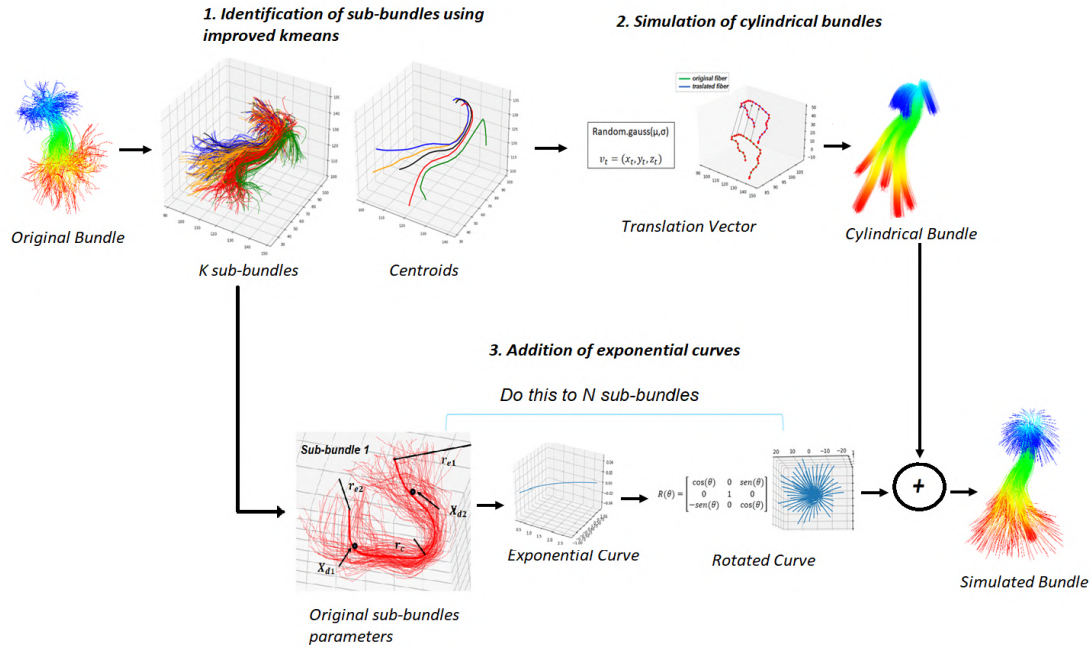


Figura 5.1.1: Esquema general del algoritmo de simulación utilizando curvas exponenciales. Los pasos principales del método incluyen: 1) Identificación de sub-fascículos utilizando Kmeans mejorado, 2) Simulación de fascículos cilíndricos, 3) Adición de curvas exponenciales. Los principales parámetros de un sub-fascículo son: r_{e1} y r_{e2} , definidos como los radios de los primeros y segundos extremos, r_c como el radio central, y x_{d1} y x_{d2} como la posición de inicio de la dispersión para cada extremo.

5.1.1.1. Identificación de sub-fascículos utilizando el algoritmo Kmeans mejorado

El algoritmo Kmeans tradicional utiliza un muestreo aleatorio para inicializar los centroides, lo que no garantiza una distribución espacial óptima. Esto puede ocasionar agrupamientos ineficientes debido a la cercanía excesiva entre algunos centroides iniciales. Para superar esta limitación, el algoritmo Kmeans++ introduce un método mejorado de inicialización, seleccionando de forma iterativa centroides iniciales que están más alejados entre sí, comenzando desde un punto aleatorio [33]. Esto ayuda a evitar divisiones incorrectas y mejora la calidad de los *clusters*.

Sin embargo, Kmeans++ puede seleccionar puntos atípicos como centroides

iniciales, lo que podría afectar la calidad de los resultados. Para mitigar este problema, se introduce una etapa adicional denominada retracción de puntos. Esta técnica ajusta los centroides iniciales en dirección al valor promedio de todos ellos, agrupándolos una fracción más cerca de la media local [54]. La Ecuación 5.1.1 describe el cálculo de la retracción:

$$Q_i = IC_i * (1 - r) + IC_m * r, \quad (5.1.1)$$

Donde Q_i es el centroide retraído, IC_i son los centroides iniciales seleccionados mediante Kmeans++, IC_m es el promedio de todos los centroides iniciales, y r es la tasa de retracción. Esta indica el porcentaje que deben acercarse los puntos a la media, donde 0 indica que los puntos permanecen en su lugar y 1 que todos los puntos se sobreponen a la media. Este ajuste mejora la distribución espacial de los centroides, reduciendo la influencia de puntos atípicos. En este trabajo se utiliza una razón de retracción del 5% como fue propuesto por Goicovich et al. (2021).

Para aplicar el algoritmo a un modelo de fascículos, se calcula una matriz de distancias entre fibras, utilizando la distancia Euclidiana máxima [39]. Posteriormente, se aplica el algoritmo Kmeans mejorado (Kmeans++ con retracción para la inicialización de los centroides) sobre esta matriz, obteniendo un conjunto de clusters de fibras y sus respectivos centroides. Cada *cluster* resultante representa un sub-fascículo identificado en el modelo.

5.1.1.2. Generación de fascículos cilíndricos

Para generar el fascículo cilíndrico, se utilizan uno o más centroides. Estos se desplazan a lo largo de cada uno de sus ejes coordenados mediante un vector de traslación, el cual se calcula empleando una distribución normal basada en los parámetros de media (μ) y varianza (σ) definidos por el usuario. La cantidad de

desplazamientos que se aplican a los centroides está determinada por el porcentaje que representa cada *cluster* dentro del fascículo y por el número total de fibras que se desea simular.

5.1.1.3. Adición de las curvas exponenciales

Los fascículos en general tienen una forma dispersa en los extremos. Debido a esto, una representación tubular no es suficiente. Para representar mejor la forma de los fascículos, se utilizan curvas exponenciales que se añaden a las fibras del fascículo cilíndrico. Las curvas exponenciales se modelan para representar diferentes radios en cada extremo del sub-fascículo (r_{e1} y r_{e2}) y en el centro (r_c). Además, se incluyen las posiciones de inicio de la dispersión para cada extremo (x_{d1} y x_{d2}). Los parámetros se introducen en el modelo utilizando una función exponencial natural e^u .

Dado que los extremos de los subfascículos son diferentes, cada subfascículo se divide en dos secciones, una a cada extremo, utilizando el centro como punto de referencia. Para modelar estas secciones, se emplean dos curvas exponenciales cuya forma depende del radio de cada extremo. En el caso de los fascículos basados en el atlas de DWM, los radios de los extremos y del centro de cada sub-fascículo se calculan promediando las distancias desde el centroide hasta las demás fibras. La ecuación general de la curva exponencial (C_i) para un extremo $i=1,2$ se muestra en la ecuación 5.1.2.

$$C_i = F_a(r_{ei}e^u + r_c), \quad (5.1.2)$$

Donde r_{ei} representa el radio de cada extremo, r_c representa el radio central u es un vector de muestras de puntos en el intervalo $[1, -1]$. El número de muestras depende de la posición inicial de la dispersión. F_a es un factor de ajuste utilizado

para compensar la diferencia en el número de fibras entre el sub-fascículo original y el simulado. Este factor permite obtener el grado de apertura del sub-fascículo en los extremos. F_a se define como la razón entre el número de fibras simuladas N_{fs} y el número de fibras del fascículo original N_{fo} , como se muestra en la ecuación 5.1.3.

$$F_a = \frac{N_{fs}}{N_{fo}} \quad (5.1.3)$$

Si $N_{fs} > N_{fo}$ entonces $F_a > 1$, lo que aumenta la amplitud de la curva exponencial, simulando una mayor apertura en los extremos del fascículo. En caso contrario se reduce la amplitud de la curva y el extremo tendrá un menor radio que el fascículo original.

La curva resultante es rotada haciendo uso de una matriz de rotación con un ángulo de rotación $\theta_i = \frac{i}{N_f} \cdot 360^\circ$ donde $i = 1, 2, \dots, N_f$. De esta forma, es rotada de manera uniforme alrededor de un círculo, completando una rotación de 360° .

Posteriormente la curva rotada es adicionada a cada una de las fibras del fascículo tubular. Una vez conformado el fascículo es posible agregar ruido aleatorio en los extremos de las fibras usando un vector aleatorio para obtener la apariencia dispersa de los extremos.

Las posiciones de dispersión (x_{d1} y x_{d2}) son los puntos desde los cuales se deben agregar las exponenciales para ambas secciones de fibras, respectivamente. Estos parámetros son configurables por el usuario y definen cómo se distribuyen los extremos del sub-fascículo. Al utilizar un modelo de fascículo, x_{d1} and x_{d2} se obtienen de los sub-fascículos originales. Para obtener esta posición, se calcula para cada fibra el ángulo de apertura de cada punto de fibra con respecto al centroide del sub-fascículo. Luego, se obtienen los ángulos promedios para todas las fibras del sub-fascículo. Finalmente, se utiliza el índice del punto que presenta

el mayor cambio en el ángulo medio como la posición de inicio de la dispersión.

A continuación, a cada fibra del fascículo cilíndrico se le agrega una curva correspondiente para simular la dispersión de fibras distribuidas alrededor del centroide del sub-fascículo. La curva calculada se rota utilizando una matriz de rotación y luego se suma a la fibra respectiva. Una vez que se forma un sub-fascículo, es posible agregar ruido al final de las fibras utilizando un vector aleatorio con valores en el intervalo $[0, 1]$ para obtener una apariencia irregular en los extremos. Este proceso se aplica a cada sub-fascículo.

Se utilizaron tres fascículos con diferentes formas del atlas DWM para validar el simulador de fascículos. Estos fascículos son el fascículo uncinado, el tracto corticoespinal y el fascículo arqueado. Además, para cada fascículo, se usaron cinco sub-fascículos y el mismo número de fibras que los fascículos originales.

Se realizó una comparación cuantitativa calculando la distancia inter-fascículo y el porcentaje de intersección entre los fascículos originales y los simulados [80]. Se computó una matriz de distancias inter-fascículo utilizando la distancia Euclidiana máxima entre puntos correspondientes [39] entre cada fibra del fascículo original y cada fibra de los fascículos simulados correspondientes. Para cada fila, se selecciona la distancia mínima, que corresponde a la fibra más cercana en el fascículo simulado. La distancia inter-fascículo se calcula como la distancia media entre las fibras más cercanas de ambos fascículos. Además, se computa la desviación estándar de las distancias entre las fibras más cercanas.

Además, el porcentaje de intersección se calcula como el porcentaje de fibras en el fascículo original que tienen una fibra similar en el fascículo simulado [80]. Dos fibras se consideran similares cuando la distancia entre ellas es menor que un umbral definido en 10 mm [80, 72].

5.1.2. Fascículos simulados en todo el cerebro para la validación del algoritmo de clustering QB

Como ejemplo de una aplicación, se generaron cuatro conjuntos de datos de todo el cerebro con fascículos simulados para validar un algoritmo de agrupamiento (*QuickBundles*). Los conjuntos de datos están compuestos por fascículos simulados, en los cuales los centroides fueron seleccionados de la tractografía de todo el cerebro de un sujeto de la base de datos ARCHI [84]. Un diagrama general de los pasos se muestra en la Figura 5.1.2.

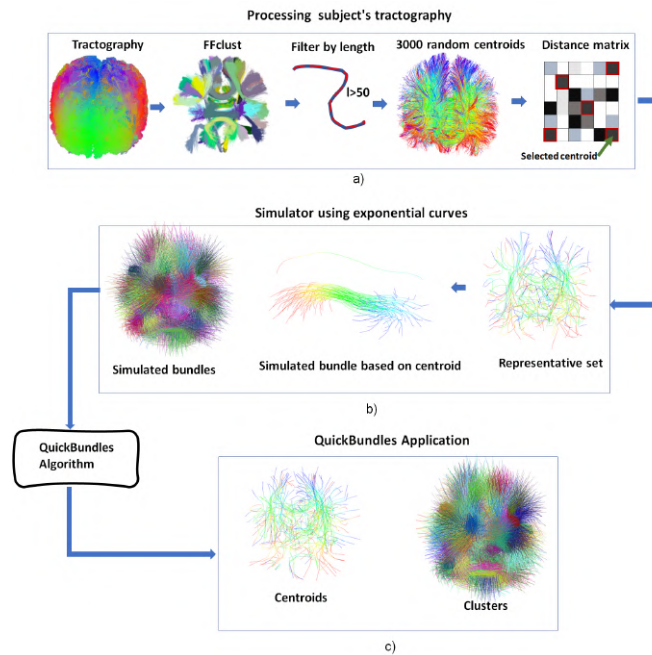


Figura 5.1.2: Esquema general de la generación de conjuntos de datos de fascículos simulados para validar el algoritmo *QuickBundles* (QB). a) El procesamiento de la tractografía del sujeto incluye la reducción de dimensionalidad con el algoritmo FFClust, el filtrado por longitud de fibra y la selección de las fibras más distantes de un conjunto aleatorio de fibras. b) El simulador utilizando curvas exponenciales muestra el conjunto representativo de centroides, un fascículo simulado basado en un centroide del conjunto y todos los fascículos simulados. c) La aplicación de *QuickBundles* muestra los centroides y los *clusters* obtenidos al aplicar QB al conjunto de datos simulado.

Primero, se reduce la dimensionalidad de los datos y se obtiene un conjunto de

centroides intra-sujeto representativos ($\approx 49,000$) para la tractografía de un sujeto al aplicar el algoritmo de *clustering* FFClust [96]. A continuación, para descartar fibras cortas y ruidosas, los centroides se filtraron por longitud, manteniendo solo aquellos con una longitud superior a 50 mm. Finalmente, se utilizó un subconjunto de 3000 centroides intra-sujeto aleatorios para seleccionar un conjunto de las fibras más representativas. La selección se realizó calculando una matriz de distancias entre fibras y encontrando aquellas que están más alejadas entre sí. Se crearon cuatro conjuntos de datos de fascículos simulados utilizando como entrada los centroides intra-sujeto seleccionados. Se trabajó con dos conjuntos de centroides intra-sujeto, uno con 100 y el otro con 500 centroides. Estos centroides se emplearon como los centroides de los fascículos simulados.

Para simular un fascículo, se utilizó su centroide y se especificó un número aleatorio de fibras variando en el rango de 50 a 300 fibras. Los componentes del vector de traslación se determinaron aleatoriamente siguiendo una distribución normal con una media $\mu = 0$ y una desviación estándar σ , que varía entre 1.0 y 2.0 mm. El radio central r_c varió en el rango de 4 a 8 mm. Dado que se trabaja con fibras de 21 puntos, se establecieron las posiciones de dispersión, x_{d1} y x_{d2} , en 10 (comenzando desde el punto central). Además, se agregó ruido aleatorio a cada punto de la fibra en el rango $[0, 1]$ para dar naturalidad a las curvas. Finalmente, se crearon dos conjuntos de datos para cada conjunto de centroides (con 100 y 500 fascículos), utilizando diferentes rangos de radios de extremo (r_{e1} y r_{e2}): de 5 a 10 mm para simular fascículos con dispersión moderada, y de 10 a 15 mm para generar fascículos más dispersos.

El algoritmo de *clustering* de fibras *QuickBundles* se aplicó a los cuatro conjuntos de datos de fascículos simulados. Su rendimiento fue evaluado para dos umbrales de distancia, 10 mm y 15 mm, comúnmente utilizados en el estado del arte [29, 30, 96]. Se calcularon dos medidas para evaluar la calidad de los *clusters* obtenidos según

el *ground truth*: el número de *clusters* recuperados al 100 % en un solo *cluster* y la división media de *clusters*, es decir, el número promedio de *clusters* en los que se dividió cada fascículo simulado.

5.2. Resultados

5.2.1. Validación del simulador usando un DWM atlas

La validación del simulador se realizó mediante la simulación de tres fascículos del atlas (el fascículo uncinado, el tracto corticoespinal y el fascículo arqueado). Figura 5.2.1 muestra que los fascículos simulados con extremos dispersos tienen una apariencia más realista y una mejor similitud con el atlas de fascículos en comparación con los fascículos cilíndricos. Los resultados de la distancia interfascículos y el porcentaje de intersección se presentan en la Tabla 5.2.1.

Fascículo	Distancia Media (mm) \pm STD	% Intersección
Uncinado	9,65 \pm 2,63	65,15
Arqueado	15,02 \pm 4,23	21,44
Tracto Corticoespinal	8,17 \pm 2,14	82,03

Tabla 5.2.1: Distancia media en mm, desviación estándar y porcentaje de intersección entre los fascículos simulados y los fascículos del atlas.

Los resultados cuantitativos muestran un buen desempeño del simulador, con distancias interfascículo generalmente bajas y altos porcentajes de intersección para los fascículos simulados. El tracto corticoespinal, un fascículo largo y regular, obtuvo los mejores resultados con aproximadamente 8 mm de distancia interfascículo y un porcentaje de intersección del 82 %. En cambio, se obtuvieron valores más bajos para el fascículo arqueado (15 mm y 22 %, respectivamente), probablemente debido a su forma y estructura complejas.

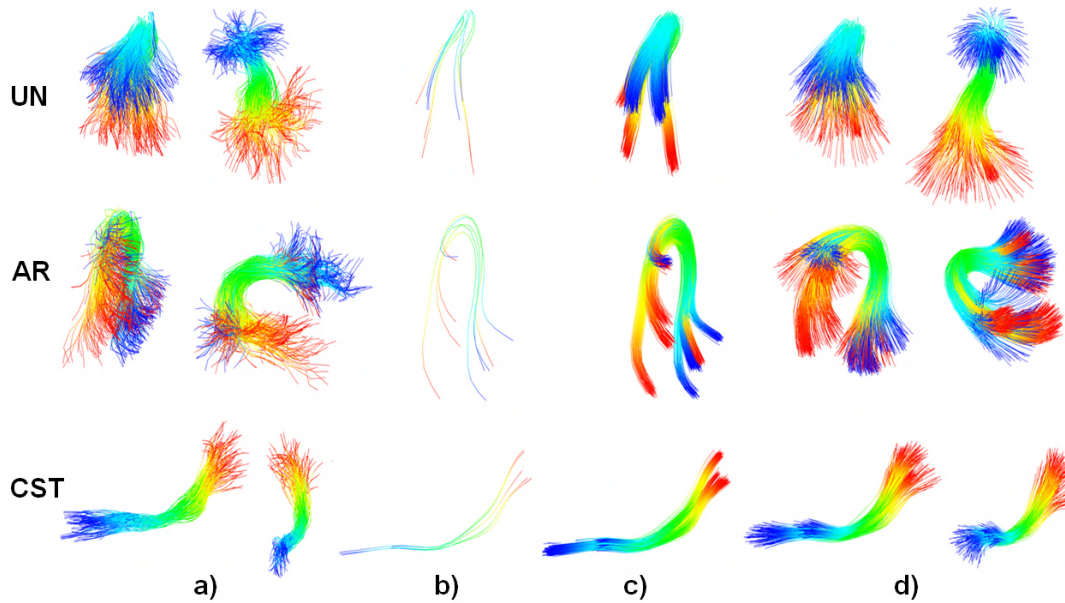


Figura 5.2.1: Resultados finales de la simulación para los tres fascículos del atlas, utilizando cinco sub-fascículos y el mismo número de fibras que los fascículos originales: el fascículo uncinado, el tracto corticoespinal y el fascículo arqueado. a) Dos vistas del atlas de fascículos originales, b) Centroides obtenidos del Kmeans mejorado, c) Fascículos cilíndricos obtenidos mediante la traslación de los centroides, y d) Dos vistas del fascículo simulado utilizando curvas exponenciales

5.2.2. Fascículos simulados en todo el cerebro para la validación del algoritmo de clustering QB

La Figura 5.2.2 muestra los dos conjuntos de centroides utilizados para simular los conjuntos de datos de fascículos de todo el cerebro, así como ejemplos de fascículos y los cuatro conjuntos de datos de fascículos simulados. Estos se obtuvieron utilizando las dos configuraciones para el parámetro de los radios (5-10 mm y 10-15 mm). La Tabla 5.2.2 presenta los resultados obtenidos para los *clusters* recuperados al 100% y la media de división de los *clusters* utilizando diferentes configuraciones de los radios. A partir de la tabla, se puede ver que el rendimiento de *QuickBundles* es generalmente mejor para los conjuntos de datos con 100 centroides, lo cual es esperado, ya que los conjuntos de datos con más de 500

fascículos presentan menor separación entre fascículos.

Además, se observó que el uso del algoritmo de *clustering QuickBundles* con un umbral de 15 mm proporciona mejores resultados para todos los conjuntos de datos, lo que es coherente con los datos simulados. Sin embargo, dado que el MDF es una distancia promedio entre puntos correspondientes de dos fibras, usar un umbral de 10 mm es demasiado pequeño para subdividir correctamente los fascículos de entrada.

En cuanto a los radios en los extremos, se puede observar que valores más altos de estos parámetros (en el rango de 10-15 mm) conducen a más *clusters* de QB (con un porcentaje muy bajo de recuperación de fascículos) y una mayor división media de *clusters*. Esto probablemente se deba a la mayor distancia entre las fibras en los extremos de los fascículos simulados. El peor caso fue, como era de esperarse, el conjunto de 500 centroides con radios grandes en los extremos, donde QB no recuperó ningún fascículo original para ambos umbrales.

Parámetros del simulador		QuickBundles thr=10 mm			QuickBundles thr=15 mm		
Original Centroides	Radios externos Rango (mm)	QB Clusters	QB Fascículos recuperados (100%)	Media de división de los clusters	QB Clusters	QB Fascículos recuperados (100%)	Media de división de los clusters
100	5-10	295	25 (25%)	2.82	123	35 (35%)	1.65
	10-15	636	1 (1%)	6.60	522	6 (6%)	5.36
500	5-10	1180	37 (7%)	3.75	540	95 (19%)	2.05
	10-15	2893	0	12.20	1922	0	8.67

Tabla 5.2.2: Resultados de rendimiento del algoritmo QuickBundles (QB) con dos umbrales de distancia diferentes (10 y 15 mm) para los cuatro conjuntos de datos de fascículos simulados.

5.3. Discusión

Este capítulo presentó una primera aproximación a un novedoso algoritmo de simulación de fascículos. Actualmente, no existe en el estado del arte una solución similar para superar parcialmente la falta de *ground truth* en los métodos de *clustering* de fibras. De esta manera, se ofrece una herramienta para validar este

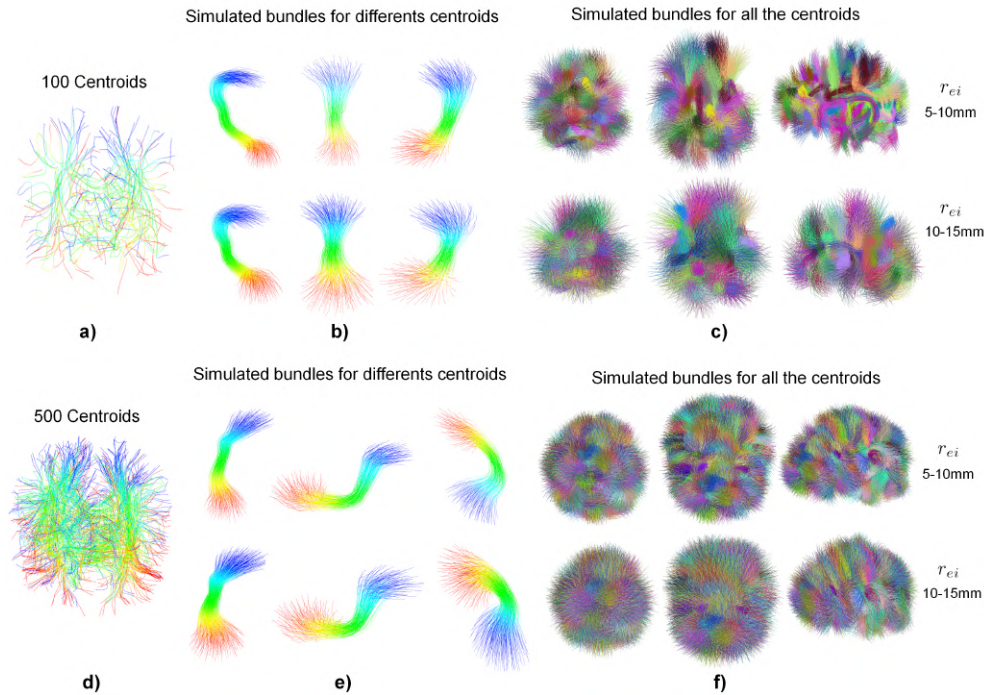


Figura 5.2.2: Fascículos simulados para los dos conjuntos de centroides. a) Conjunto de 100 centroides, b) Tres fascículos simulados para radios en los extremos (r_{e1} y r_{e2}) en el intervalo de 5-10 mm (**arriba**) y 10-15 mm (**abajo**) para el conjunto de 100 centroides, c) Conjuntos de datos de fascículos simulados para el conjunto de 100 centroides y los dos rangos de radios en los extremos. d), e) y f): Las mismas visualizaciones que en a), b) y c), respectivamente, pero para el conjunto de 500 centroides.

tipo de métodos, lo cual es muy importante para aportar nuevas perspectivas sobre la estructura de la materia blanca del cerebro.

El simulador propuesto utiliza parámetros de forma del fascículo, como los radios en los extremos, el radio central y la posición inicial de la dispersión, que se incorporan al algoritmo utilizando curvas exponenciales para garantizar la apertura en los extremos del fascículo. Como resultado, el simulador puede generar fascículos con un porcentaje de intersección moderado a alto al simular fascículos del atlas. La aplicabilidad del simulador se probó mediante la generación de cuatro conjuntos de datos de fascículos con diferentes números de fascículos y radios en los extremos, que se utilizaron para evaluar el rendimiento del algoritmo *QuickBundles*. Se

demonstró que los resultados del *clustering* son coherentes con los parámetros de entrada utilizados tanto para el simulador como para el algoritmo el *clustering*, lo que demuestra la utilidad del marco propuesto para generar datos simulados de fascículos.

Es relevante destacar que los parámetros de entrada fueron seleccionados únicamente para demostrar la aplicabilidad del simulador y no como una validación del algoritmo QB. Se requieren análisis adicionales para definir un rango adecuado de parámetros de configuración que permita validar de manera más efectiva los algoritmos de *clustering*.

Aunque el simulador ofrece resultados prometedores, presenta algunas limitaciones. Actualmente, las curvas se generan mediante la traslación de los centroides, lo que reduce la variabilidad en la forma de las fibras simuladas. Además, la adición de curvas exponenciales rotadas en los extremos restringe la diversidad de los fascículos generados. Para abordar estas limitaciones, se podría integrar curvas capaces de representar una mayor variedad de formas complejas, lo que permitiría una simulación más realista de las fibras cerebrales. Una opción viable sería el uso de curvas *splines*, una técnica previamente empleada en la representación de fibras [11].

A pesar de sus limitaciones, el simulador basado en curvas exponenciales demuestra el potencial de este enfoque para la validación de algoritmos de *clustering*, destacando su utilidad y relevancia para la comunidad científica.

Capítulo 6

Simulador de fascículos usando curvas spline

En el capítulo anterior, se presentó una primera aproximación a un simulador de fascículos de fibras. Sin embargo, el uso de curvas exponenciales en los extremos limita la representación a fibras que no presenten múltiples curvaturas a lo largo de su trayecto, lo que restringe también las formas de los fascículos que pueden ser simulados. Este capítulo introduce la utilización de curvas *spline* para la representación de las fibras. Dado que estas son curvas polinómicas, permiten simular una mayor variedad de formas de las fibras y de una manera más cercana a la forma real.

6.1. Métodos

En esta sección se explica la metodología utilizada para desarrollar el simulador de fascículos de fibras, utilizando curvas *spline* y parámetros de forma de fascículos. Para evaluar la robustez del simulador, se implementa un proceso de validación [72] mediante la simulación de fascículos de un atlas de fascículos de materia

blanca profunda [39].

Para comparar los fascículos simulados con los fascículos del atlas de referencia, se emplearon dos métricas de similitud: una métrica de distancia entre fascículos y el porcentaje de intersección entre fascículos. A diferencia del Capítulo 5, donde sólo se simularon tres fascículos para el proceso de validación, en este se simulan 28 fascículos del atlas para evaluar el comportamiento del simulador para una mayor variedad de formas de fascículos.

Por último, se utiliza el simulador para generar conjuntos de datos simulados de tractografía de todo el cerebro que sirvan como *ground truth* para evaluar el rendimiento de diferentes algoritmos de clustering de fibras. En este caso se evalúan dos algoritmos de *clustering* de fibras: *QuickBundles* y *FFClust*. El desempeño de ambos algoritmos en la recuperación de los *clusters* del *ground truth* se evaluó para diferentes umbrales de distancia, en tres conjuntos de datos con 100, 500 y 1000 fascículos, y diferentes permutaciones de los datos simulados.

6.1.1. Simulador de fascículos usando curvas spline

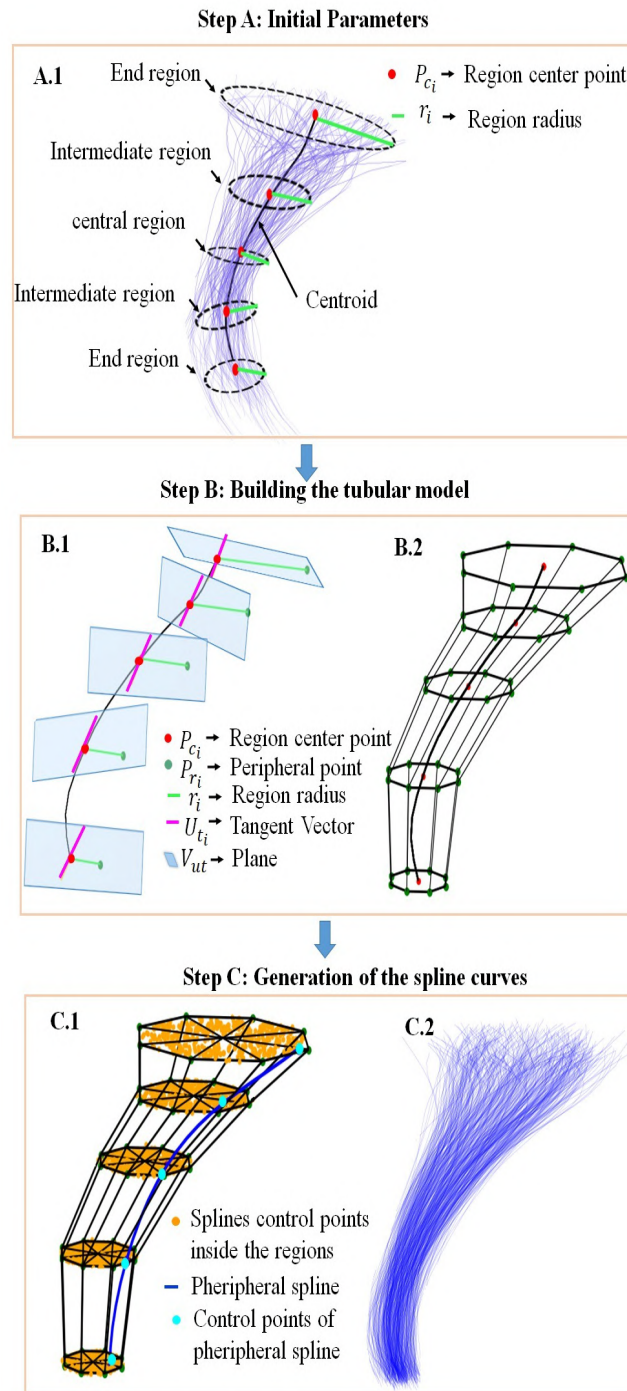
La simulación de los fascículos se realizó considerando un centroide y la división del fascículo en cinco regiones transversales: las dos regiones extremas, la región central y dos regiones intermedias. Estas regiones fueron seleccionadas ya que los radios de las regiones externa y central de un fascículo son características importantes a tener en cuenta a la hora de describir la forma del fascículo [102]. En esta investigación, dado que solo se dispone de información del centroide, también se consideran los radios de dos regiones transversales intermedias. Los parámetros clave para el proceso de simulación del fascículo fueron el centroide, que proporciona una descripción aproximada de la trayectoria del fascículo, los radios de cada región y el número de fibras a simular.

La metodología consiste en construir un modelo tubular para la simulación del fascículo. Se utilizan círculos en cada sección transversal seleccionada, centrados en sus respectivos puntos centroides. Para crear un círculo, se selecciona un punto inicial en su periferia y se gira siete veces alrededor de su punto central, con ángulos múltiplos de 45° con el objetivo de generar 8 sectores circulares. Esto es un parámetro que puede modificarse.

La clave para obtener una curva realista fue utilizar puntos dentro de cada círculo como puntos de control para construir las *splines*. Se utilizaron *splines* de cuarto orden para garantizar la suavidad y continuidad de las curvas finales del fascículo simulado. La Figura 6.1.1 muestra el esquema general del proceso, y a continuación se describe cada etapa.

6.1.1.1. Paso A: Inicialización de parámetros

La Figura 6.1.1 (A.1) muestra la división del fascículos en 5 regiones transversales. Los parámetros de entrada para cada fascículo del algoritmo de simulación son el centroide del fascículo, el número de fibras, y los radios de las cinco regiones transversales r_i , donde el subíndice i en cada parámetro representa el número de la región, de 1 a 5. Cada radio representa la distancia desde la fibra más lejana del fascículo hasta el punto correspondiente del centroide. Para garantizar la dispersión real existente en los extremos de los fascículos, se añade opcionalmente ruido gaussiano a los cinco primeros puntos de cada extremo, definido por la media y la varianza gaussianas como parámetros de entrada. Para definir los centros de las cinco regiones transversales del fascículo, el algoritmo selecciona cinco puntos específicos a lo largo del centroide (P_{ci}). Dado que se trabaja con fibras compuestas por 21 puntos, se utilizaron puntos con índices 0, 3, 10, 17 y 20. Estos puntos demostraron ser representativos de la forma del fascículo en trabajos anteriores [47, 96].



6.1.1.2. Paso B: Construcción del modelo tubular

Para construir el modelo tubular que representa la forma del fascículo, se generan cinco círculos alrededor de los puntos centroides seleccionados, definiendo así las cinco regiones. Para crear cada círculo, se elige un punto inicial en su periferia (P_{ri}) y se rota en torno a su respectivo centro P_{ci} utilizando siete ángulos de rotación diferentes θ . Estos ángulos son múltiplos de 45° , cubriendo el intervalo de 0° a 315° , lo que resulta en la división del círculo en ocho sectores de 45° cada uno.

La Figura 6.1.1 (B.1) ilustra cómo obtener el punto periférico inicial de cada círculo (P_{ri}). En primer lugar, se determina la dirección del centroide del fascículo en cada punto central P_{ci} , calculando el vector tangente (\vec{U}_{ti}) a la curva en ese punto. A continuación, para definir el punto periférico inicial de cada sección transversal P_{ri} , se selecciona un punto a una distancia r_i de P_{ci} , siguiendo la dirección del vector $\overrightarrow{P_{ci}P_{ri}}$ (perpendicular al vector \vec{U}_{ti}), y perteneciente al plano V_{ut} que contiene a P_{ci} definido por \vec{U}_{ti} . Para asegurar que todos los puntos están en posiciones correspondientes, se alinean con respecto al primer punto del círculo.

Para la rotación de P_{ri} alrededor de P_{ci} , se construye una matriz de rotación 3x3 [79, 90] basada en el eje de rotación correspondiente $\vec{U}_{ti} = (u_{xi}, u_{yi}, u_{zi})$ y el ángulo de rotación θ , como se muestra en la Ec. 6.1.1. Se selecciona \vec{U}_{ti} como eje de rotación para cada P_{ri} , lo que garantiza que la rotación representa fielmente la dirección del centroide del fascículo en cada región. Figura 6.1.1 (B.2) muestra el modelo tubular obtenido tras la rotación inicial de los puntos periféricos de cada círculo, obteniendo un conjunto de ocho puntos que definen los círculos.

$$\mathbf{R} = \begin{bmatrix} u_{x_i}^2(1 - \cos \theta) + \cos \theta & u_{x_i}u_{y_i}(1 - \cos \theta) - u_{z_i}\sin \theta & u_{x_i}u_{z_i}(1 - \cos \theta) + u_{y_i}\sin \theta \\ u_{y_i}u_{x_i}(1 - \cos \theta) + u_{z_i}\sin \theta & u_{y_i}^2(1 - \cos \theta) + \cos \theta & u_{y_i}u_{z_i}(1 - \cos \theta) - u_{x_i}\sin \theta \\ u_{z_i}u_{x_i}(1 - \cos \theta) - u_{y_i}\sin \theta & u_{z_i}u_{y_i}(1 - \cos \theta) + u_{x_i}\sin \theta & u_{z_i}^2(1 - \cos \theta) + \cos \theta \end{bmatrix} \quad (6.1.1)$$

6.1.1.3. Paso C: Generación de las curvas spline

El paso anterior genera un modelo tubular, dividido en ocho sectores circulares a lo largo de las cinco secciones transversales. Para cada sector circular, el algoritmo genera puntos siguiendo una distribución aleatoria uniforme, como se muestra en la Figura 6.1.1 (C.1). Esta disposición de los puntos genera curvas distribuidas por todo el modelo tubular. Para la representación de las fibras cerebrales se utilizan *splines* de cuarto orden ya que este orden es una buena aproximación para describir la complejidad de la forma de las fibras. Nótese que para el seguimiento de fibras se han utilizado *splines* de tercer orden [51, 99], pero se prefirió utilizar un orden superior para garantizar la suavidad de las curvas.

Para construir los *splines* de cuarto orden, se necesitan al menos cinco puntos de control [93, 10]. Las curvas *spline* se generan sector por sector, seleccionando un punto dentro del mismo sector en cada círculo. La Figura 6.1.1 (C.1) ilustra la creación de una *spline* (en azul) junto con sus correspondientes puntos de control (en cian).

Para añadir dispersión a los extremos de la curva, se añade opcionalmente ruido gaussiano a los puntos de ambos extremos de la *spline*. Para las pruebas, en cada extremo del fascículo, se añade ruido a los cinco primeros puntos de cada curva, con una desviación estándar, σ , que varía en el rango de 2.5-3.5 mm. Como se utilizaron fibras con 21 puntos, los índices de estos puntos son 0, 1, 2, 3, 4 para el primer extremo, y 16, 17, 18, 19, 20 para el otro extremo. La adición de ruido ayuda a obtener una curva más realista con una variabilidad natural. El ruido añadido puede modificarse en función del grado de dispersión deseado. En el ejemplo, un conjunto de todos los *splines* obtenidos utilizando el modelo tubular se muestra en la Figura 6.1.1 (C.2).

6.1.2. Validación del simulador de fascículos mediante un DWM atlas

El simulador se validó utilizando un atlas de fascículos DWM [39]. Como parámetros para la simulación se utilizaron el centroide, el radio de cada sección y el número de fibras correspondientes a cada fascículo del atlas original. Para determinar el centroide de cada fascículo del atlas, primero se alinearon las orientaciones de las fibras del fascículo con respecto a una fibra de referencia.

La alineación de las fibras dentro de un fascículo es necesaria, ya que estas pueden almacenarse en memoria en cualquiera de los dos órdenes posibles con respecto a una fibra de referencia: orden directo u orden inverso. Para lograr esta alineación, se analiza primero la orientación de cada fibra en relación con la fibra de referencia, comparando la distancia euclidiana entre sus extremos. Si la distancia entre los extremos opuestos es mayor, la fibra se reorienta invirtiendo el orden de todos sus puntos en la memoria.

La fibra de referencia se seleccionó de la siguiente manera: primero, se eligieron las fibras con una longitud superior a 50 mm para descartar aquellas más cortas, que suelen ser más ruidosas. Luego, de entre las fibras restantes, se seleccionó la que tuviera la distancia promedio mínima respecto a todas las demás fibras. A continuación, se calculó el centroide como la media de las fibras del fascículo. El radio de cada círculo se calculó como la media de las distancias euclidianas entre el punto central del círculo y todos los puntos de las fibras correspondientes del fascículo. La orientación del fascículo se define en función del centroide. A partir de los 21 puntos que definen el centroide, se determinó una aproximación *spline*, obteniéndose los coeficientes de la ecuación de la curva. Con ellos, para obtener los vectores tangentes, se utilizó la ecuación de la primera derivada de la curva y se calculó su valor en cada punto representativo del centroide. Los vectores tangentes

contienen la orientación del centroide en esos puntos.

Para medir el grado de similitud de los fascículos simulados con los fascículos originales del atlas, y también para comparar los resultados con el simulador anterior basado en curvas exponenciales (Ver Capítulo 5), se utilizaron dos métricas de similitud: la distancia entre fascículos y el porcentaje de intersección entre fascículos. La distancia entre fascículos representa la distancia media entre las fibras más cercanas de dos fascículos. El porcentaje de intersección entre un par de fascículos se calcula como el porcentaje de fibras similares entre los dos fascículos. Dos fibras se consideran similares si la distancia entre ellas es inferior a 10 mm [80, 72]. El cálculo de estas métricas se basa en el uso de una matriz de distancias, donde las filas de la matriz representan la distancia entre las fibras del fascículo simulado y el fascículo del atlas. Como medida de la distancia entre las fibras, se utilizó la distancia euclidiana máxima entre los puntos correspondientes [80]. Esta medida considera las dos direcciones posibles de las fibras, como se muestra en la Ec. 6.1.2, donde a_i y b_i son los índices de los puntos en las fibras A y B, respectivamente. N_p corresponde al número de puntos de la fibra.

$$d_{ME}(A, B) = \min(\max_i ||a_i - b_i||, \max_i ||a_i - b_{(N_p-i)}||) \quad (6.1.2)$$

Para comprobar la viabilidad de utilizar un modelo tubular para representar los fascículos, se realizó un breve análisis de la forma de las regiones de los fascículos del atlas, modelándolos con secciones transversales elípticas. La Figura A.1.1 muestra un ejemplo de una representación de una región intermedia de sección transversal de un fascículo usando círculos y elipses. La Figura A.1.2 muestra el fascículo representado usando el modelo tubular y el modelo preliminar basado en secciones transversales elípticas. A continuación, se calculó la razón entre los radios menor y mayor para las cinco regiones transversales de los fascículos del atlas

analizados, así como la razón media de los radios. Un valor cercano a 1.0 indica que ambos radios son muy similares y, en consecuencia, la forma se aproximará más a un círculo. Como se muestra en la Tabla A.1.1, las relaciones de radios son, en general, superiores a 0.7, con una razón de radios media total de 0.82, lo que indica que es razonable como primera aproximación utilizar un modelo tubular con secciones transversales circulares para representar fascículos DWM tubulares.

Para la validación del simulador, se excluyeron 8 fascículos con forma de lámina o *sheet-like* del análisis, ya que el algoritmo está diseñado para simular fascículos tubulares. Estos fascículos corresponden al cuerpo calloso y al cíngulo corto, los cuales presentan una morfología completamente diferente debido a que abarcan áreas más extensas de la corteza. Además, suelen subdividirse en múltiples fascículos mediante algoritmos de *clustering* exploratorio. Por lo tanto, se simularon 28 fascículos, incluidas las versiones izquierda y derecha de los fascículos del fascículo arqueado, las fibras largas del cíngulo, el tracto corticoespinal, el fórnix, el fascículo fronto-occipital inferior, el fascículo longitudinal inferior, las radiaciones talámicas y el fascículo uncinado.

6.1.3. Fascículos simulados en todo el cerebro para la validación de los algoritmos de clustering

En este estudio, se generaron conjuntos de datos de fascículos simulados de todo el cerebro que sirven de base para la validación de algoritmos de *clustering*. La metodología utilizada incluye los siguientes parámetros: los centroides de los fascículos, los radios de los círculos que representan las regiones transversales, el número de fibras de cada fascículo y los parámetros de ruido gaussiano. El objetivo es mostrar la practicidad del simulador propuesto para evaluar el comportamiento de dos algoritmos de *clustering*, *QuickBundles* (QB) [29] y *Fast Fiber Clustering* (FFClust) [96].

Los centroides de los fascículos se seleccionaron a partir de la tractografía de todo el cerebro de un sujeto de la base de datos ARCHI [84]. Esta tractografía comprende una colección de fibras, cada una con 21 puntos en \mathbb{R}^3 . En primer lugar, se filtraron las fibras cortas y ruidosas seleccionando únicamente los centroides con una longitud superior a 50 mm. A continuación, se eligieron centroides a una distancia euclidiana máxima de 10 mm del conjunto de fibras filtradas para que actuaran como centroides de fascículos simulados. Se trabajó con tres conjuntos de centroides compuestos por 100, 500 y 1000 centroides.

Para seleccionar el rango de parámetros de simulación para los radios de los círculos y el número de fibras, se realizó un estudio de parámetros basado en los fascículos del atlas de fascículos DWM [41]. Donde para cada fascículo del atlas se calcularon los radios promedio de cada una de las regiones y el número total de fibras. Como resultado, la configuración de parámetros de la simulación consta de los siguientes rangos para cada radio: r_1 y r_5 : 8-10 mm (círculos extremos), r_2 y r_4 : 6-8 mm (círculos intermedios), r_3 : 5-7 mm (círculo central). Para definir los radios, el algoritmo funciona del siguiente modo. En primer lugar, los radios externos se seleccionan aleatoriamente y por separado dentro de su intervalo definido. Luego, los radios intermedios se seleccionan aleatoriamente y por separado dentro de su intervalo correspondiente, pero por debajo de sus radios externos. Por último, el radio central se selecciona aleatoriamente dentro de su rango, pero inferior a los radios de las regiones intermedias. Este procedimiento pretende evitar formas irregulares al construir el modelo tubular. A continuación, se añadió un ruido gaussiano, con un σ variable en el rango de 2.5-3.5 mm. Estos parámetros del ruido se seleccionaron de manera visual; pero pueden ser cambiados por el usuario si se desea una mayor o menor dispersión en los extremos. El número de fibras se eligió entre 50 y 300. Todas las selecciones aleatorias se basan en una distribución normal.

Se generaron tres conjuntos de datos simulados con 100, 500 y 1.000 fascículos de fibras. Se utilizó el formato *bundles*, ya que permite almacenar todos los fascículos en un único archivo binario (con la extensión *.bundlesdata*) y las respectivas etiquetas de los fascículos de fibras en un archivo de texto (con la extensión *.bundles*). Posteriormente, se aplicó QB y FFClust a los conjuntos de datos de tractografía simulados utilizando cuatro umbrales de distancia diferentes.

Para aplicar los algoritmos a un conjunto de datos simulados, QB solo requiere el umbral de distancia (θ_{QB}) para la agrupación (ver Sec. 2.3.1). Se utilizaron cuatro valores de θ_{QB} : 10 mm, 12 mm, 15 mm y 20 mm. En el caso de FFClust, se necesitan tres parámetros (ver Sec. 2.3.1.2): el número de *clusters* (Kp) utilizados en el MiniBatch KMeans para los diferentes puntos (Kp_{end} , Kp_{cent} y Kp_{inter}), que se obtienen mediante el método Elbow, como proponen Vázquez et al. (2020). Los parámetros restantes, que corresponden al umbral de distancia para la reasignación de los *clusters* y un umbral para la fusión de los candidatos a *cluster* (d_{Mmax}), se establecieron utilizando los umbrales de distancia $d_{Rmax} = d_{Mmax}$: 10 mm, 12 mm, 15 mm y 20 mm.

Otro punto importante que hay que tener en cuenta al evaluar el rendimiento de los algoritmos de *clustering* es el impacto que tiene en el algoritmo cambiar el orden de los elementos de entrada. Para abordar esta consideración, se realizaron cinco permutaciones de los datos de entrada y se analizaron los cambios en las métricas de *clustering*. Para realizar estas permutaciones, se utilizó el algoritmo de Fisher-Yates [25], que permuta aleatoriamente una secuencia de elementos. Empieza por el último elemento, elige un índice al azar y lo intercambia con el elemento actual. Repite este proceso hacia atrás en la secuencia. Este método garantiza permutaciones aleatorias uniformes y es eficiente en términos de tiempo. Por último, se evaluó la calidad de los *clusters* obtenidos por el algoritmo comparándolos con el *ground truth*. Se utilizaron métricas estandarizadas, como

accuracy geometric, *precision*, *recall* y *F-measure*, para evaluar la calidad de los *clusters* obtenidos por los algoritmos de *clustering*. Además, muchos estudios de investigación han utilizado la medida *Overlap Score* (OS) propuesta por [Nepusz et al. \(2012\)](#). El objetivo de OS es evaluar el grado de concordancia entre los *clusters* predichos por cualquier método de acuerdo con los *clusters* del *ground truth* [68, 46, 42]. La Ec. 6.1.3 define el OS entre C_p y C_g correspondientes al *cluster* predicho por el algoritmo y el *cluster* real de *ground truth*, respectivamente.

$$OS(C_p, C_g) = \frac{|C_p \cap C_g|^2}{|C_p||C_g|} \quad (6.1.3)$$

Precision evalúa la proporción de *clusters* correctamente identificados y predichos por el algoritmo entre todos los *clusters* predichos. Por el contrario, *recall* mide la fracción de *clusters* reales que el algoritmo predice correctamente. *F-measure* es la media armónica de *precision* y *recall*. Precision, recall y F-measure se definen del siguiente modo:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6.1.4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6.1.5)$$

$$\text{F-measure} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6.1.6)$$

donde TP es el número de *True Positives*, que representa el número de *clusters* que coinciden con los *clusters* de *ground truth* con un OS igual o superior a un umbral. En este trabajo se utiliza un OS de 0.8 como fue sugerido por [Nepusz et al. \(2012\)](#). FP representa el número de *False Positives* y es la diferencia entre el número de

clusters predichos por el algoritmo y el número de TP. FN o *False Negatives* se refiere al número de *clusters* que no coinciden con los *clusters* predichos por el algoritmo. Como se ha observado, los valores TP, FP y FN se basan en el *Overlap Score*.

También se utilizaron otras dos medidas: *Geometric Accuracy* (Acc) y *Maximum Matching Ratio* (MMR) [7, 68, 42] porque aportan perspectivas adicionales que complementan la evaluación del algoritmo de *clustering*. En primer lugar, Acc representa la media geométrica de *Clustering Wise Sensitivity* (S_n) y *Positive Predictive Value* (PPV). S_n indica la habilidad del algoritmo para identificar fibras dentro del *ground truth*, específicamente en lo que respecta a la cobertura. En cambio, el PPV denota la probabilidad de TP en los *clusters* predichos por el algoritmo. Entonces S_n , PPV y Acc pueden definirse como se muestra en las ecuaciones 6.1.7, 6.1.8 y 6.1.9.

$$S_n = \frac{\sum_{i=1}^n \max_j \{t_{ij}\}}{\sum_{i=1}^n N_i} \quad (6.1.7)$$

$$PPV = \frac{\sum_{j=1}^m \max_i t_{ij}}{\sum_{j=1}^m T_j} \quad (6.1.8)$$

$$Acc = \sqrt{S_n \times PPV} \quad (6.1.9)$$

donde t_{ij} es el número de fibras en común entre el *cluster* i_{th} del *ground truth* y el *cluster* j_{th} predicho por el algoritmo. El número de *clusters* en el *ground truth* es n y m es el número de *clusters* predichos por el algoritmo. N_i es el número de fibras presentes en el i_{th} *clusters*, y $T_j = \sum_{i=1}^n \max_j t_{ij}$.

En segundo lugar, *Maximum Matching Ratio* mide la proporción de coincidencia

de los *clusters* predichos por el algoritmo con los *clusters* del *ground truth*. MMR penaliza los casos en los que un *cluster* del *ground truth* se divide en múltiples partes en el conjunto predicho. Utiliza un grafo bipartito con aristas ponderadas para medir la calidad del conjunto predicho en relación con el conjunto del *ground truth*. El puntaje MMR refleja la suma de los pesos de las aristas dividida por el número de *clusters* del *ground truth*. Las aristas se ponderan en función de la puntuación de solapamiento entre los *clusters*. MMR se calcula sumando todos los valores de SO de los *clusters* predichos por TP divididos por el número de *clusters* del *ground truth*.

6.2. Resultados

Todos los experimentos se realizaron en una computadora equipada con una CPU Intel(R) Core(TM) i7-9700 a 3.0 GHz y 32 GB de RAM, con Windows 10 Pro (64 bits). Se describen los resultados del simulador de fascículos utilizando 28 fascículos DWM, seguidos de los experimentos de validación de los algoritmos de *clustering* de fibras.

6.2.1. Validación del simulador usando un DWM atlas

Para validar el simulador propuesto, se generaron 28 fascículos del atlas DWM con diferentes formas y número de fibras. Para comparar los resultados del simulador con los fascículos del atlas se utilizaron dos métricas: la distancia entre fascículos y el porcentaje de intersección entre fascículos. Estas métricas, basadas en la distancia euclidiana máxima entre fibras (Ec. 6.1.2, ver Sección, 6.1.2), permitieron evaluar la precisión de la representación de la forma del fascículo. La Tabla 6.2.1 presenta los resultados detallados de ambas métricas.

La Figura 6.2.1 muestra tres resultados de simulación correspondientes a fascículos

Atlas Bundles	Previous simulator (I)		Previous simulator (II)		Proposed simulator	
	ID \pm STD (mm)	IP (%)	ID \pm STD (mm)	IP (%)	ID \pm STD (mm)	IP (%)
Anterior Left Arcuate	10.5 \pm 4.4	61.2	10.6 \pm 4.2	59.1	8.2 \pm 1.1	96.4
Anterior Right Arcuate	10.6 \pm 3.9	60.7	10.9 \pm 4.2	58.6	8.2 \pm 1.0	97.9
Left Arcuate	15.0 \pm 4.2	21.4	17.3 \pm 5.4	7.9	10.5 \pm 2.2	44.8
Right Arcuate	13.9 \pm 4.8	25.2	15.9 \pm 4.7	8.5	10.8 \pm 2.2	39.9
Posterior Left Arcuate	8.3 \pm 2.9	81.3	8.7 \pm 3.7	68.9	6.6 \pm 1.0	100.0
Posterior Right Arcuate	8.3 \pm 2.9	77.0	8.9 \pm 3.5	67.2	6.7 \pm 0.9	100.0
Left Cingulum Temporal Fibers	7.5 \pm 2.5	81.5	9.8 \pm 4.6	57.3	7.9 \pm 1.6	86.4
Right Cingulum Temporal Fibers	7.9 \pm 2.7	81.5	10.1 \pm 4.9	57.3	7.2 \pm 1.4	89.0
Left Corticospinal Tract	8.2 \pm 2.1	82.0	13 \pm 7.2	47.4	8.1 \pm 1.6	87.3
Right Corticospinal Tract	6.8 \pm 3.1	85.6	12.8 \pm 7.4	52.4	8.0 \pm 1.6	87.5
Left Fornix	9.1 \pm 3.6	68.5	13.6 \pm 5.4	31.4	10.6 \pm 1.8	32.9
Right Fornix	9.2 \pm 3.8	68.6	13.8 \pm 5.8	32.9	10.8 \pm 1.7	30.0
Left Inferior Fronto-occipital	14.2 \pm 6.4	31.9	16.1 \pm 7.7	8.9	11.9 \pm 1.7	13.2
Right Inferior Fronto-occipital	14.1 \pm 6.1	27.11	16.5 \pm 7.9	8.2	11.5 \pm 1.5	15.5
Left Inferior Longitudinal	11.0 \pm 4.8	45.4	16.6 \pm 7.8	20.8	10.3 \pm 1.8	48.2
Right Inferior Longitudinal	11.9 \pm 5.4	46.4	17.0 \pm 5.4	20.8	10.2 \pm 1.8	49.6
Left Frontal Thalamic Radiations	9.4 \pm 3.8	62.3	10.5 \pm 4.2	55.9	7.9 \pm 1.7	88.5
Right Frontal Thalamic Radiations	10.1 \pm 5.2	56.5	10.9 \pm 5.8	50.1	8.1 \pm 1.8	85.4
Left Superior Motor Thalamic Radiations	7.1 \pm 2.6	86.2	10.4 \pm 6.3	59.7	7.1 \pm 1.1	99.3
Right Superior Motor Thalamic Radiations	7.2 \pm 3.1	84.8	9.4 \pm 5	66.4	7.2 \pm 1.3	98.2
Left Occipital Thalamic Radiations	9.5 \pm 4.5	71.1	11.5 \pm 7.1	55.4	7.6 \pm 1.6	90.4
Right Occipital Thalamic Radiations	10.4 \pm 5.0	64.0	11.7 \pm 6.8	55.4	7.7 \pm 1.5	92.8
Left Superior Parietal Thalamic Radiations	7.3 \pm 2.9	81.9	8.9 \pm 5.1	73.2	6.3 \pm 1.1	100.0
Right Superior Parietal Thalamic Radiations	7.5 \pm 3.2	79.8	9.4 \pm 5.8	66.4	7.5 \pm 1.2	100.0
Left Temporal Thalamic	8.5 \pm 3.3	75.4	10.9 \pm 5.5	50.8	7.2 \pm 1.6	93.8
Right Temporal Thalamic	8.5 \pm 2.7	75.4	10.9 \pm 4.8	50.8	7.2 \pm 1.4	95.4
Left Uncinate	9.6 \pm 2.6	65.2	11.1 \pm 5.2	54.8	8.4 \pm 1.2	90.5
Right Uncinate	9.3 \pm 3.9	67.3	10.3 \pm 4.1	56.3	8.6 \pm 1.2	88.6
Mean Values	9.7 \pm 3.8	64.8	11.9 \pm5.7	46.5	8.5 \pm 1.5	76.5

Tabla 6.2.1: Resultados de métricas entre los fascículos simulados y los fascículos del atlas para el simulador usando curvas spline y el simulador basado en curvas exponenciales [72]: Simulador anterior I (simulación con cinco centroides), Simulador anterior II (simulación con un centroide), Distancia entre fascículos \pm Desviación estándar (mm) (ID \pm STD) y Porcentaje de intersección entre fascículos (%) (IP).

del atlas con diferentes formas: el fascículo uncinado izquierdo, el fascículo longitudinal inferior izquierdo y el fascículo fronto-occipital inferior. El primer fascículo simulado, mostrado en la Figura 6.2.1 (A), alcanzó una buena calidad de representación con un porcentaje de intersección de fascículos del 90.5% y una distancia entre fascículos de 8.4 ± 1.2 mm. La Figura 6.2.1 (B) muestra un resultado de calidad media, con un porcentaje de intersección del 48.2% y una distancia entre fascículos de 10.3 ± 1.8 mm. La Figura 6.2.1 (C) muestra un ejemplo de resultado de baja calidad, con un porcentaje de intersección del 13.2% y una distancia entre fascículos de 11.9 ± 1.7 mm. El resultado de calidad media o baja se debe a fascículos cuyos extremos están muy dispersos y tienen radios variables. Por lo tanto, una representación que utilice un único centroide y un radio medio de las fibras al centroide sólo puede capturar parcialmente la forma del fascículo. Figura A.1.3 y Figura A.1.4 del Apéndice muestran los fascículos

simulados restantes del atlas.

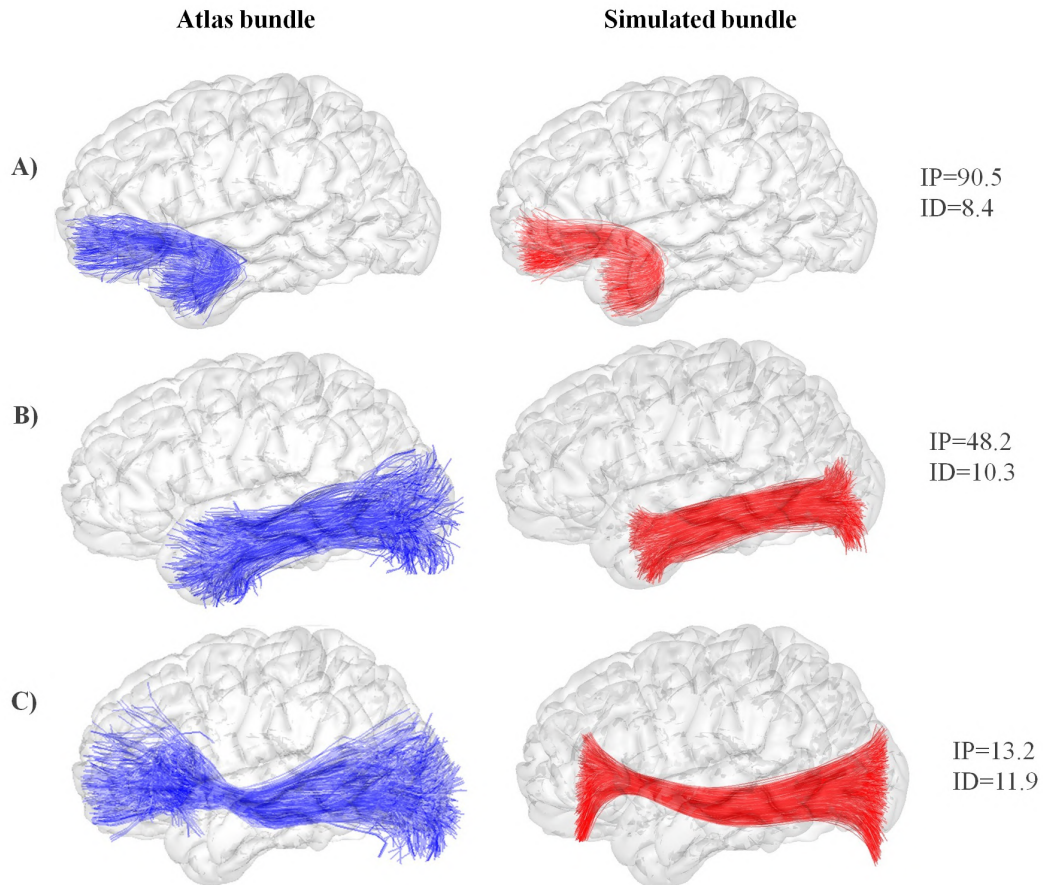


Figura 6.2.1: Ejemplos de simulación de fascículos basados en el atlas de fascículos DWM [39]. Los fascículos originales del atlas se muestran a la izquierda (azul), y sus correspondientes fascículos simulados a la derecha (rojo). A) El fascículo uncinado izquierdo, B) El longitudinal inferior izquierdo y C) El fronto-occipital inferior. IP: Porcentaje de intersección entre fascículos (%) ID: Distancia entre fascículos (mm).

Se realizaron también simulaciones utilizando el simulador basado en curvas exponenciales (ver Capítulo 5). Al igual que en el capítulo anterior, se simularon los fascículos del atlas descomponiéndolos en cinco subfascículos. Sin embargo, esta estrategia utiliza cinco centroides; por lo tanto, para tener una comparación más justa de los modelos de fascículos, también se simularon fascículos con curvas exponenciales utilizando un solo centroide. La Tabla 6.2.1 presenta los resultados de la comparación.

Los resultados muestran un buen rendimiento del simulador propuesto, con una distancia media interfascículo (ID) (y desviación estándar) de 8.5 ± 1.5 mm y un porcentaje medio de intersección (PI) del 76.5 %. El método supera al simulador anterior basado en cinco centroides, con métricas medias de 9.7 ± 3.8 mm y 64.8 %, respectivamente, y basado en un solo centroide, con métricas medias de 11.9 ± 5.7 mm y 46.5 %, respectivamente. Estos resultados muestran que el simulador puede reproducir con éxito la mayoría de los fascículos. En concreto, 20 de los 28 fascículos simulados tenían un porcentaje de intersección superior al 85 % con el correspondiente fascículo del atlas. No obstante, algunos fascículos, como el fascículo fronto-occipital inferior izquierdo y derecho, no se simularon correctamente. Estos fascículos tienen formas complejas con extremos separados, lo que dificulta su simulación utilizando la información proporcionada por un solo centroide. La Tabla 6.2.1 muestra que el simulador anterior que utilizaba cinco subfascículos obtenía mejores porcentajes de intersección para estos fascículos específicos, pero cuando utilizaba sólo un centroide, presentaba un rendimiento inferior. Es importante señalar que el cálculo del centroide del fascículo también afecta a los resultados de la simulación. En este caso, se utilizó la fórmula de centroide más común, basada en la media de los puntos correspondientes al fascículo. Sin embargo, este enfoque al obtener el centroide mediante el promedio de los puntos y naturalmente las fibras no tienen el mismo largo ni la misma curvatura, esto da lugar a centroides más cortos que se ubican cerca del centro del fascículo y que no abarcan toda su extensión.

Adicionalmente, la Figura A.1.1 y la Tabla A.1.1 muestran los resultados de simulación obtenidos utilizando el modelo preliminar basado en secciones transversales elípticas, en comparación con el modelo tubular propuesto. La Figura A.1.5 muestra los tres fascículos mostrados en la Figura 6.2.1 para el modelo preliminar elíptico. En la Tabla A.1.1 se puede observar observar que sólo

se encontraron mejoras en IP e ID en ocho fascículos (tractos espinales corticales izquierdo y derecho, fórnices izquierdo y derecho, radiaciones talámicas occipitales izquierda y derecha, y fascículo uncinado izquierdo y derecho), con resultados comparables o inferiores para los fascículos restantes. Por lo tanto, los trabajos futuros podrían mejorar este modelo, encontrando mejores estrategias para definir los radios de las elipses.

6.2.2. Validación de algoritmos de clustering de fibras usando el simulador

Utilizando el modelo tubular, se simularon tres conjuntos de datos de todo el cerebro con 100, 500 y 1000 fascículos, que sirvieron de base para evaluar el rendimiento de dos algoritmos de *clustering* de fibras. La Tabla 6.2.2 muestra las principales características del *ground truth* para los diferentes números de fascículos, incluyendo el número de centroides utilizados para la simulación y la distancia mínima y media entre ellos. La Figura A.1.6 muestra los conjuntos de centroides. En la Tabla 6.2.2 también se enumeran el número mínimo y máximo de fibras por fascículo, el número total de fibras del *ground truth* y el número de fascículos que se cruzan. Una vista más detallada de cada conjunto de datos de fascículos simulados se muestra en el la Figura A.1.7, Figura A.1.8 y Figura A.1.9.

Caracterización del Ground Truth			
Número de centroides	100	500	1000
Distancia mínima entre centroides (mm)	13.23	10.16	10.01
Distancia promedio entre centroides (mm)	90.03	91.69	90.81
Mínimo número de fibras por fascículo	50	50	50
Máximo número de fibras por fascículo	300	300	300
Número total de fibras	19,052	87,728	176,300
Número total de fascículos cruzados	15	274	619

Tabla 6.2.2: Caracterización del *ground truth*: El número de centroides utilizados para la simulación, la distancia mínima y media entre centroides, el número mínimo y máximo de fibras en los fascículos, el número total de fibras en el *ground truth* y el número de fascículos con cruzamientos.

Dado que los fascículos simulados contienen las fibras generadas alrededor de sus centroides, algunos fascículos pueden contener fibras que se acercan mucho a las fibras de otros fascículos. Estos fascículos se denominan fascículos cruzados. Se definen los fascículos cruzados como aquellos que contienen fibras que se encuentran a una distancia inferior a 10 mm de las fibras de otros fascículos. La Figura 6.2.2 muestra diferentes tipos de fascículos cruzados en el *ground truth*.

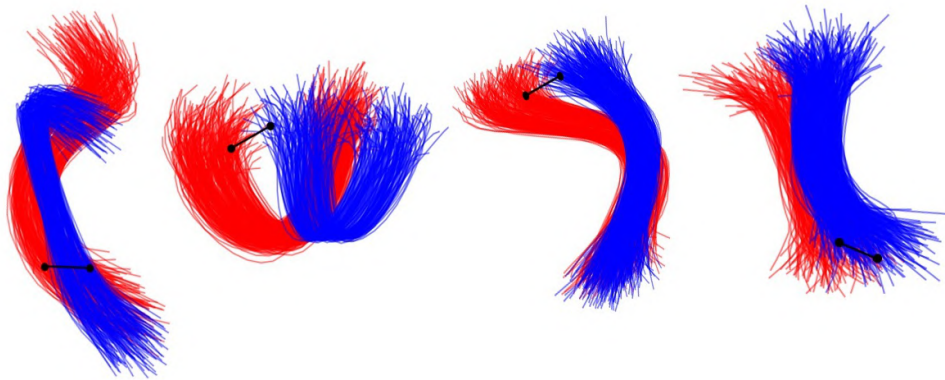


Figura 6.2.2: Ejemplos de diferentes tipos de cruzamientos existentes entre dos fascículos en los conjuntos de datos del *ground truth*. El fascículo rojo y el azul representan dos fascículos diferentes del *ground truth*. La línea negra representa un marcador de distancia de 10 mm.

Se aplicaron los algoritmos QB y FFClust con diferentes umbrales de distancia de 10, 12, 15 y 20 mm a estos conjuntos de datos simulados. Como se mencionó en la Sección 6.1.3, para el algoritmo FFClust se utilizó el método del codo para establecer el número de *clusters* por punto de fibra que es otro de sus parámetros, junto a los umbrales de distancia. La Figura A.1.10 muestra las curvas del método del codo para los tres conjuntos de datos reales.

La Figura 6.2.3 muestra los *clusters* obtenidos por ambos algoritmos para los diferentes umbrales de distancia. Como se observa, existe un mayor cruce entre los fascículos de los conjuntos de datos II y III (500 y 1000 fascículos), correspondiente a la información mostrada en la Tabla 6.2.2, debido al mayor número de fascículos utilizados. Además, el comportamiento de ambos algoritmos es bastante similar,

aumentando el tamaño de los *clusters* a medida que se incrementa el umbral de distancia.

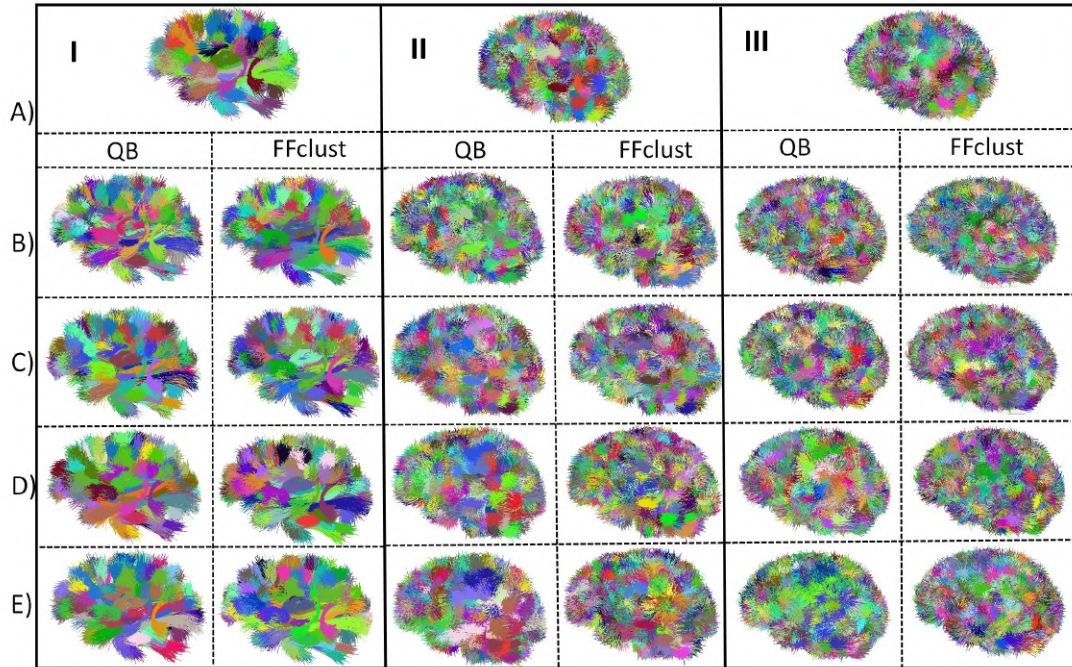


Figura 6.2.3: Resultados obtenidos del *clustering* de fibras de QB y FFClust para los conjuntos de datos de tractografía simulados con 100 fascículos (I), 500 fascículos (II) y 1000 fascículos (III): A) Fascículos simulados originales. *Clusters* obtenidos por QB y FFClust para diferentes umbrales: B) 10 mm, C) 12 mm, D) 15 mm, E) 20 mm.

Para evaluar la calidad de los *clusters* obtenidos por los algoritmos, se calcularon las cinco medidas descritas anteriormente: *Accuracy* (Acc), *Precision*, *Recall*, *F-measure* y *Maximum Matching Ratio* (MMR). Los resultados de las distintas medidas para ambos algoritmos pueden verse en las Tabla 6.2.3 y Tabla 6.2.4 para los conjuntos de datos de tractografía simulada con 100, 500 y 1000 fascículos, respectivamente.

Las Tabla 6.2.3 y 6.2.4 muestran el número de *clusters* predichos por los algoritmos. Las Tabla A.1.2 y A.1.3 muestran los valores de TP, FP, FN, PPV y Sn. En general, ambos algoritmos obtuvieron buenos resultados a medida que aumentaban los umbrales de distancia. Mientras que QB tiende a dividir menos, FFClust tiende a

fusionar menos, comportamientos esperados para ambos algoritmos y consistentes para el diferente número de *clusters*. Como QB tiene como único parámetro de entrada el umbral de distancia (θ_{QB}), su comportamiento sólo depende de este valor. A medida que aumenta el umbral, se crean menos *clusters* y se añaden a los *clusters* fibras situadas a mayor distancia de los centroides de los *clusters*. Si θ_{QB} es demasiado grande, los grupos de fibras con formas diferentes se agruparán formando *clusters* más grandes y gruesos. Por otro lado, FFClust tiene más parámetros que influyen en los resultados finales. Los valores del número de *clusters* (Kp) para los diferentes puntos en el primer paso (Kp_{end} , Kp_{inter} , y Kp_{cent}) son los más importantes, ya que influyen en el tamaño de los *clusters* de puntos. Un mayor número de *clusters* de puntos producirá *clusters* de fibras preliminares más pequeños. Los dos umbrales de distancia también influyen en los resultados del *clustering*, reasignando los *clusters* pequeños a *clusters* grandes (d_{Rmax}) y fusionando los *clusters* que comparten el *cluster* de puntos central (d_{Mmax}). Por lo tanto, los umbrales de distancia más altos también producirán *clusters* más grandes y gruesos, pero en menor medida que QB, por lo que es menos probable que fusione fibras con formas diferentes, pero, por el contrario, es más probable que subdivida más los *clusters*.

Original <i>clusters</i>	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
Algorithm clusters	168	111	91	76	748	468	328	199	1432	890	572	283
Accuracy	0.87	0.95	0.93	0.87	0.81	0.86	0.81	0.66	0.79	0.81	0.73	0.56
Precision	0.31	0.72	0.85	0.75	0.21	0.52	0.55	0.36	0.21	0.48	0.48	0.30
Recall	0.51	0.79	0.77	0.57	0.31	0.48	0.36	0.14	0.29	0.42	0.28	0.08
F-Measure	0.38	0.75	0.81	0.65	0.25	0.50	0.43	0.20	0.24	0.45	0.35	0.13
Maximum Matching Ratio	0.49	0.78	0.77	0.56	0.30	0.47	0.35	0.14	0.28	0.40	0.27	0.08

Tabla 6.2.3: Valores de las métricas para evaluar el rendimiento del algoritmo QB para los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se resaltan las métricas para el umbral con el mejor rendimiento.

Como se observa en Tabla 6.2.3, Tabla 6.2.4 y en Figura 6.2.4, QB obtuvo su mejor rendimiento con un umbral de 12 mm para todos los conjuntos de datos

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
Algorithm clusters	176	155	126	117	1193	978	797	688	2722	2171	1719	1527
Accuracy	0.93	0.94	0.95	0.93	0.79	0.82	0.82	0.81	0.73	0.76	0.78	0.76
Precision	0.43	0.51	0.66	0.67	0.11	0.18	0.26	0.25	0.06	0.12	0.18	0.18
Recall	0.76	0.79	0.83	0.78	0.27	0.35	0.41	0.34	0.16	0.25	0.32	0.27
F-Measure	0.55	0.62	0.73	0.72	0.16	0.24	0.32	0.29	0.09	0.16	0.23	0.21
Maximum Matching Ratio	0.73	0.76	0.81	0.76	0.24	0.33	0.39	0.33	0.15	0.23	0.30	0.26

Tabla 6.2.4: Valores de las métricas para evaluar el rendimiento del algoritmo FFClust para los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se resaltan las métricas para el umbral con el mejor rendimiento.

reales, mientras que FFClust obtuvo el mejor rendimiento con un umbral de 15 mm. A medida que aumenta el número de fascículos, disminuyen las medidas de calidad de los *clusters*, lo que es de esperar debido al aumento de los cruces de fascículos. Estos cruces pueden confundir a los algoritmos de *clustering*.

El valor máximo de las medidas se obtuvo para el *ground truth* de 100 fascículos para ambos algoritmos, alcanzando un *Accuracy* de 0.95 en ambos casos. Además, QB logra una mayor *Accuracy* que FFClust en los conjuntos de datos más grandes. Los valores de *Accuracy* disminuyen a medida que aumentan los conjuntos de datos para ambos algoritmos, y QB es más sensible al umbral de distancia que FFClust en todos los conjuntos de datos. Estos valores se logran porque ambos algoritmos presentan una buena proporción de *True Positives* (TP), como se observa en Tabla A.1.2 y Tabla A.1.3. De manera similar, la *Precision* alcanza valores cercanos para ambos algoritmos, con 0.72 para QB y 0.66 para FFClust en el conjunto de datos más pequeño. Sin embargo, los valores de precisión se ven más afectados en FFClust con el aumento del número de fascículos, principalmente debido a un incremento en los *False Positives* (ver Tabla A.1.3). En contraste, FFClust presenta un mayor *Recall* que QB en el conjunto de datos más pequeño y menor en los conjuntos de datos más grandes.

El siguiente experimento realizado evaluó la capacidad del algoritmo para detectar correctamente fascículos cruzados. Para esto, se calculó el total de *clusters* TP

(aquellos con $OS \geq 0,8$) obtenidos por los algoritmos que correspondían a un fascículo cruzado en el *ground truth*, para cada conjunto de datos simulado y cada umbral. Luego, este conteo de TP se utilizó para calcular el porcentaje de recuperación de fascículos cruzados de ambos algoritmos como la razón entre el conteo de TP cruzados recuperados por el algoritmo y el número total de cruces reales existentes en el *ground truth*. Los resultados presentados en la Tabla 6.2.5 indican que el mejor desempeño de los algoritmos para recuperar los fascículos con cruzamientos coincide con los umbrales identificados como los más efectivos en las Tablas 6.2.3 y 6.2.4 (12 mm para QB y 15 mm para FFClust). Además, en la mayoría de los conjuntos de datos del *ground truth* y umbrales, FFClust supera a QB, lo que es consistente con las expectativas derivadas de las diferentes metodologías de agrupamiento.

Original clusters	100				500				1000			
Crossing clusters	15				274				619			
Thresholds	10	12	15	20	10	12	15	20	10	12	15	20
% crossing clusters recov QB	33.33	40.00	20.00	0.00	19.71	19.71	6.20	1.46	19.55	19.55	5.82	0.65
% crossing clusters recov FFClust	66.67	66.67	66.67	40.00	19.71	20.80	23.36	12.41	9.21	12.76	14.38	8.89

Tabla 6.2.5: Rendimiento de los algoritmos QB y FFClust en la recuperación de fascículos cruzados para diferentes umbrales (10, 12, 15 y 20 mm) y tres conjuntos de datos *ground truth* (100, 500 y 1000 fascículos): Porcentaje de *clusters* recuperados por el algoritmo (% de *clusters* cruzados recuperados).

Los valores de la F-measure, que representan un equilibrio entre las métricas de *precision* y *recall*, se ven afectados por el aumento del número de fascículos. En el caso de MMR, que proporciona una medida del grado de similitud entre los *clusters* predichos por el algoritmo y el *ground truth*, también muestra un comportamiento similar en ambos algoritmos, con FFClust superando los valores de MMR solo para el conjunto de datos más pequeño.

Las Figuras 6.2.4, 6.2.5, 6.2.6 y 6.2.7 ilustran ejemplos del rendimiento de los algoritmos en la recuperación de *clusters* específicos del *ground truth*. En la parte superior de cada figura, se presenta el fascículo original del conjunto de datos *ground truth* junto con su ubicación en el cerebro. La sección intermedia muestra

el fascículo recuperado por QB para diferentes umbrales, mientras que en la parte inferior se presenta lo mismo pero para el algoritmo FFClust. En cada figura se incluyen los valores de OS obtenidos por el fascículo del algoritmo en comparación con el *ground truth* considerando diferentes umbrales.

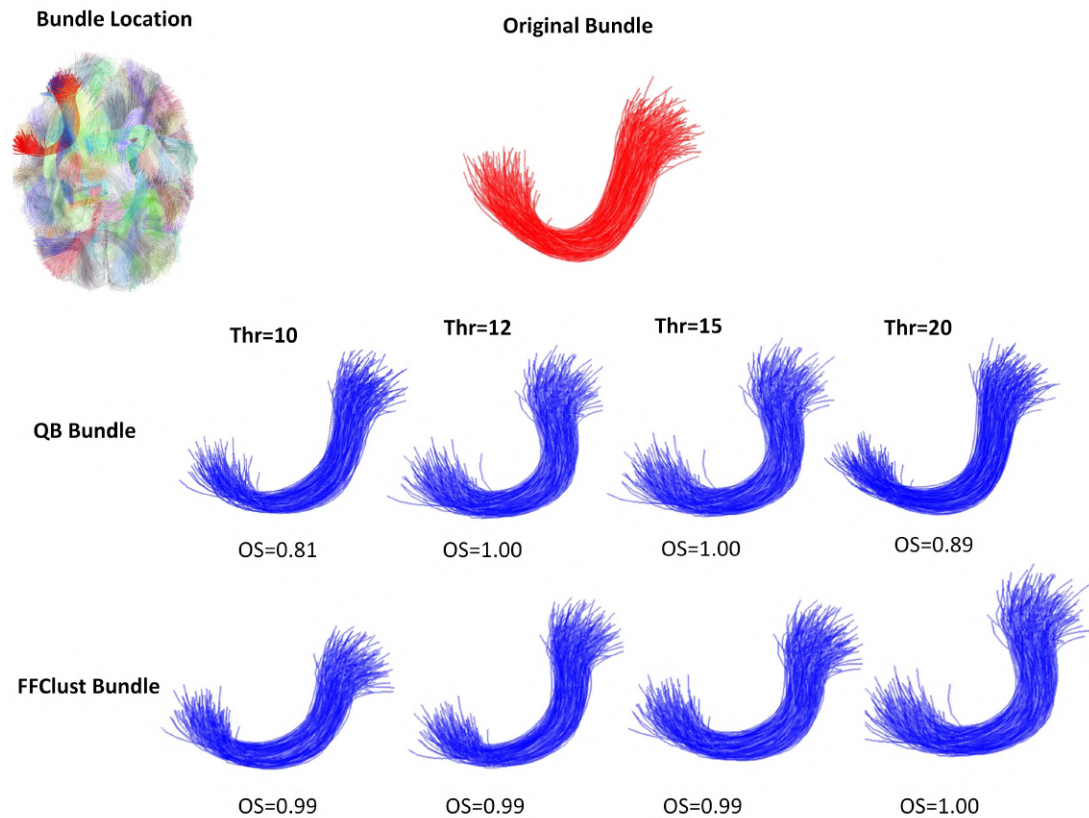


Figura 6.2.4: Un ejemplo del buen rendimiento de los algoritmos de *clustering* QB y FFClust al recuperar un *cluster* del *ground truth*. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.

La Figura 6.2.4 muestra un fascículo recuperado con éxito por ambos algoritmos, con valores de OS superiores a 0.8. En contraste, la Figura 6.2.5 muestra un rendimiento deficiente de los algoritmos al recuperar el fascículo. QB tiende a fusionar el fascículo con otros a medida que aumenta el umbral, mientras que FFClust divide el fascículo en varios fascículos, recuperando solo un pequeño porcentaje de las fibras del fascículo original. La Figura 6.2.6 presenta un ejemplo

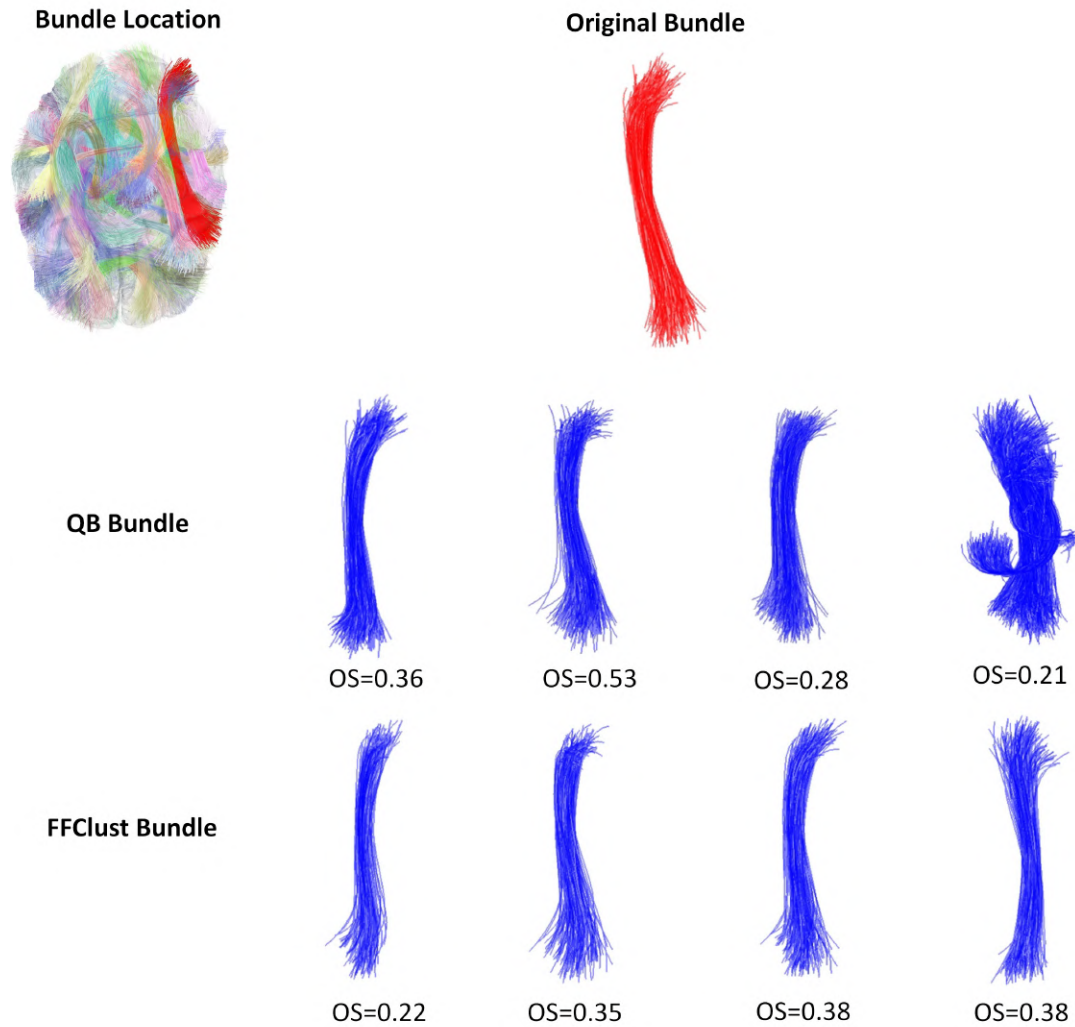


Figura 6.2.5: Un ejemplo del mal rendimiento de los algoritmos de agrupamiento QB y FFClust al recuperar un *cluster* del *ground truth*. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.

donde FFClust obtiene valores de OS más altos que QB. Esto se debe a que FFClust puede preservar el fascículo para diferentes umbrales de distancia, a diferencia de QB, que tiende a fusionar *clusters* a medida que aumenta el umbral. La explicación es que FFClust no solo depende de los umbrales de distancia, sino también del número de *clusters* de puntos en su primer paso. La Figura 6.2.7 presenta la predicción de un fascículo del *ground truth* donde QB supera a FFClust en recuperación, con un comportamiento similar para todos los umbrales. Esto se

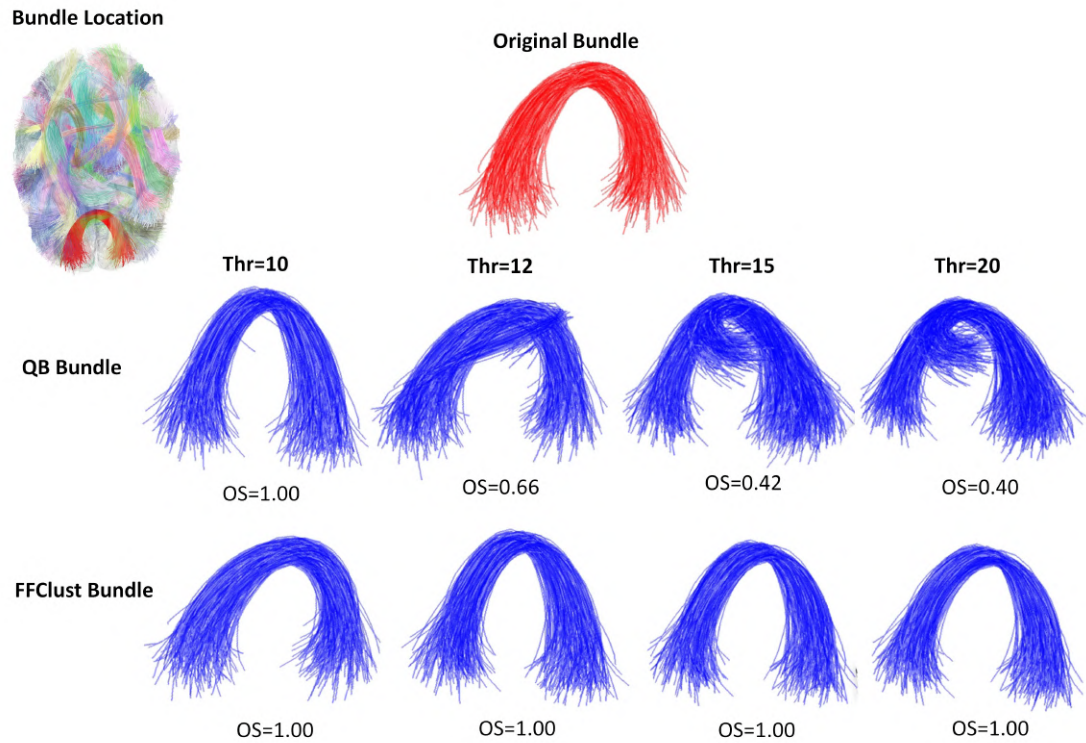


Figura 6.2.6: Un ejemplo del rendimiento de los algoritmos de *clustering* QB y FFClust al recuperar un *cluster* del *ground truth*, donde FFClust tiene un mejor desempeño que QB. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En las partes media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.

debe a la tendencia de FFClust a dividir los fascículos, especialmente cuando son grandes. Finalmente, la Figura A.1.11 y la Tabla A.1.4 del Archivo Suplementario muestran los resultados de una versión más nueva del algoritmo QB llamada QBX [30]. Este algoritmo es un 95 % más rápido que QB y los valores de las métricas para el umbral más alto son los mismos para ambos algoritmos. Sin embargo, para los umbrales de distancia restantes hay una degradación en la calidad de los resultados en comparación con los obtenidos por QB, debido a su tendencia a subdividir los fascículos en umbrales más pequeños y a crear *clusters* de menos fibras.

Otra condición importante a considerar al evaluar el rendimiento de los algoritmos

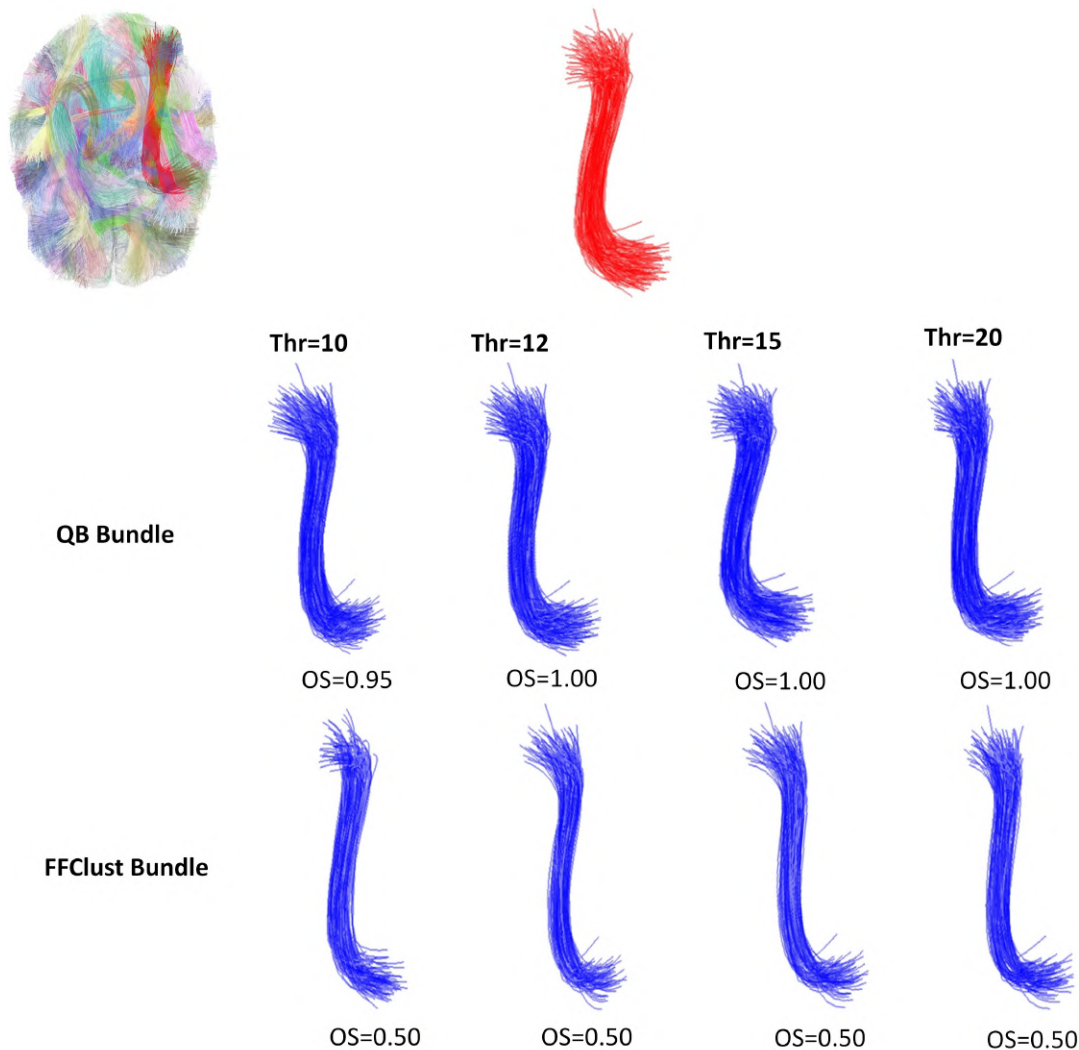


Figura 6.2.7: Un ejemplo del rendimiento de los algoritmos de *clustering* QB y FFClust al recuperar un *cluster* del *ground truth*, donde QB tiene un mejor desempeño que FFClust. La parte superior de la figura muestra el fascículo original y su ubicación en el cerebro. En la parte media e inferior, se puede observar el fascículo predicho por QB y FFClust para los diferentes umbrales y sus respectivos valores de OS.

de *clustering* es determinar la sensibilidad de los algoritmos respecto al orden de los datos de entrada. Para evaluar esto, se realizaron cinco permutaciones aleatorias de los datos de entrada para cada conjunto de datos. Se calcularon los promedios de las métricas de los algoritmos en función de las permutaciones. La Tabla 6.2.6 presenta los resultados para QB y la Tabla 6.2.7 para FFClust. Tabla A.1.5 y Tabla A.1.6 informan las variaciones de las medidas TP, FP, FN,

PPV y Sn, utilizadas para calcular las otras métricas. En general, los valores de las medidas son similares a los obtenidos en los datos originales sin permutación, con bajas desviaciones estándar. Cabe destacar que las medidas medias para todos los conjuntos de datos permutados presentaron el mejor rendimiento para los mismos umbrales que los datos de entrada originales, es decir, un umbral de 12 mm para QB y de 15 mm para FFClust. Esto muestra que los algoritmos son estables ante diferentes permutaciones de los datos de entrada.

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
Thresholds	168.33	115.33	89.67	70.67	747.67	466.67	311.00	182.00	1436.00	868.33	528.67	262.33
Algorithm clusters	± 4.04	± 3.79	± 0.58	± 2.08	± 10.21	± 3.06	± 4.36	± 4.36	± 13.08	± 8.50	± 4.51	± 3.06
Accuracy	0.87 ± 0.02	0.94 ± 0.01	0.93 ± 0.01	0.84 ± 0.02	0.79 ± 0.01	0.84 ± 0.00	0.79 ± 0.01	0.64 ± 0.01	0.77 ± 0.00	0.79 ± 0.00	0.70 ± 0.00	0.54 ± 0.00
Precision	0.28 ± 0.04	0.67 ± 0.00	0.83 ± 0.04	0.67 ± 0.04	0.21 ± 0.01	0.48 ± 0.00	0.52 ± 0.02	0.30 ± 0.02	0.21 ± 0.01	0.44 ± 0.02	0.43 ± 0.01	0.24 ± 0.01
Recall	0.46 ± 0.06	0.76 ± 0.03	0.75 ± 0.04	0.48 ± 0.04	0.31 ± 0.02	0.45 ± 0.01	0.32 ± 0.01	0.11 ± 0.01	0.30 ± 0.01	0.38 ± 0.01	0.23 ± 0.01	0.06 ± 0.00
F-Measure	0.35 ± 0.05	0.71 ± 0.01	0.79 ± 0.04	0.56 ± 0.04	0.25 ± 0.01	0.46 ± 0.00	0.40 ± 0.01	0.16 ± 0.01	0.25 ± 0.01	0.41 ± 0.02	0.30 ± 0.01	0.10 ± 0.01
Maximum Matching Ratio	0.45 ± 0.06	0.75 ± 0.02	0.74 ± 0.04	0.47 ± 0.04	0.30 ± 0.01	0.43 ± 0.00	0.32 ± 0.01	0.11 ± 0.01	0.28 ± 0.01	0.37 ± 0.01	0.22 ± 0.01	0.06 ± 0.00

Tabla 6.2.6: Rendimiento del algoritmo QB para cinco permutaciones aleatorias de los tres conjuntos de datos de tractografía simulada. La tabla muestra los valores de media y desviación estándar de todas las métricas. En azul, se destacan las métricas para el umbral con el mejor rendimiento.

6.2.3. Tiempo de ejecución y complejidad computacional del simulador

El simulador tiene tres pasos para simular un fascículo, y su complejidad temporal se basa en los pasos finales. El primer paso es la inicialización de los parámetros del modelo, que incluyen el centroide del fascículo, el número de fibras y los radios de las cinco regiones transversales. El algoritmo selecciona aleatoriamente los radios y el número de fibras de un rango de valores de entrada.

El segundo paso construye el modelo tubular, lo que requiere la generación de las

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
Thresholds	189.00	155.67	131.00	121.33	1186.33	964.33	773.33	672.67	2650.33	2150.00	1723.67	1495.00
Algorithm clusters	± 8.89	± 5.03	± 4.58	± 1.53	± 9.07	± 4.73	± 16.8	± 8.33	± 6.11	± 18.36	± 13.43	± 7.94
Accuracy	0.93 ± 0.01	0.94 ± 0.01	0.95 ± 0.01	0.94 ± 0.01	0.79 ± 0.00	0.82 ± 0.00	0.83 ± 0.00	0.81 ± 0.00	0.73 ± 0.00	0.76 ± 0.00	0.78 ± 0.00	0.76 ± 0.00
Precision	0.36 ± 0.03	0.49 ± 0.02	0.60 ± 0.04	0.62 ± 0.03	0.12 ± 0.01	0.18 ± 0.00	0.25 ± 0.02	0.26 ± 0.01	0.06 ± 0.00	0.11 ± 0.01	0.18 ± 0.01	0.18 ± 0.01
Recall	0.67 ± 0.02	0.76 ± 0.02	0.79 ± 0.04	0.75 ± 0.03	0.27 ± 0.01	0.36 ± 0.01	0.39 ± 0.02	0.35 ± 0.01	0.16 ± 0.00	0.25 ± 0.01	0.31 ± 0.01	0.27 ± 0.01
F-Measure	0.47 ± 0.03	0.59 ± 0.02	0.68 ± 0.04	0.67 ± 0.03	0.16 ± 0.01	0.24 ± 0.00	0.31 ± 0.02	0.30 ± 0.01	0.09 ± 0.00	0.16 ± 0.01	0.23 ± 0.01	0.21 ± 0.01
Maximum Matching Ratio	0.65 ± 0.02	0.74 ± 0.02	0.77 ± 0.04	0.73 ± 0.02	0.25 ± 0.01	0.33 ± 0.01	0.37 ± 0.02	0.34 ± 0.01	0.14 ± 0.00	0.23 ± 0.01	0.29 ± 0.01	0.25 ± 0.01

Tabla 6.2.7: Rendimiento del algoritmo FFClust para cinco permutaciones aleatorias del conjunto de datos de tractografía simulada. La tabla muestra los valores de media y desviación estándar de todas las métricas. En azul, se destacan las métricas para el umbral con el mejor rendimiento.

cinco regiones transversales circulares. Para cada región transversal, el algoritmo calcula puntos periféricos dentro de cada círculo. Para establecer los ocho sectores circulares, utiliza principios de geometría en 3D, incluyendo la definición de un plano, el cálculo de un producto cruzado de vectores y una matriz de rotación. Este paso es independiente del número de fibras y toma un tiempo constante.

En el tercer paso, para cada sector de cada región, el algoritmo genera puntos aleatorios uniformes de acuerdo con el número de fibras del fascículo. Para calcular cada *spline*, se tomó un punto dentro del mismo sector para cada círculo como un punto de control. Este paso es lineal en relación con el número de fibras.

Por lo tanto, la complejidad temporal para simular un fascículo es $\mathcal{O}(N)$, donde N es el número de fibras del fascículo. Así, para simular un conjunto de datos que consta de M fascículos, la complejidad temporal total es $\mathcal{O}(MN)$.

Para demostrar experimentalmente la complejidad temporal de la simulación de un fascículo, la Figura 6.2.8 muestra el tiempo de ejecución en segundos para simular un fascículo que contiene un número creciente de fibras, entre 100 y 1000 fibras, utilizando diferentes configuraciones de radios. La figura muestra que el

algoritmo tiene un tiempo de ejecución lineal para simular un fascículo. En este trabajo, se simularon tres conjuntos de datos con 100, 500 y 1000 fascículos, que se ejecutaron en 12.20, 59.84 y 122.68 segundos, respectivamente.

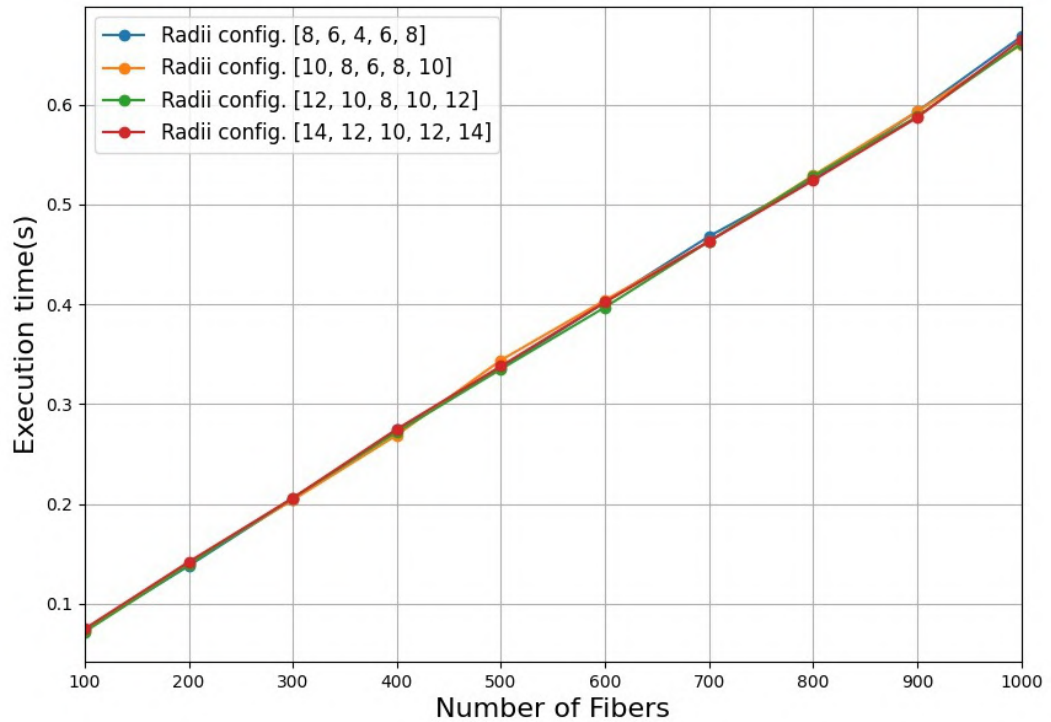


Figura 6.2.8: Tiempo de ejecución para la simulación de un fascículo que contiene un número creciente de fibras, entre 100 y 1000 fibras, utilizando diferentes configuraciones de radios. Se usan cuatro configuraciones de radios, con radios más altos para las regiones transversales externas (r_1 y r_5), un radio más bajo para la región central (r_3) y radios intermedios para las regiones r_2 y r_4 .

6.3. Discusión

Se propuso un simulador de fascículos que utiliza curvas *spline* para la representación de fibras, con el objetivo de abordar el desafío de la limitada disponibilidad de *ground truth* para los métodos de *clustering* de fibras. El simulador utiliza un centroide del fascículo y los radios de cinco regiones transversales a lo largo del centroide para construir un modelo tubular que

contiene las *splines* del fascículo simulado. Con este enfoque, se pueden simular formas de fascículos más realistas que las del estado del arte.

El simulador fue validado mediante la generación de 28 fascículos a partir de un atlas de fascículos de DWM. Como resultado, el método puede generar fascículos razonablemente similares a los fascículos del atlas, con distancias inter-fascículos generalmente bajas y altos porcentajes de intersección entre los fascículos originales y los simulados. Como era de esperar, el algoritmo funciona mejor para fascículos con formas tubulares. El simulador permite representar con éxito variaciones en los radios a lo largo de los fascículos. Sin embargo, los fascículos en forma de *sheet-like* no pueden ser representados por el método.

Los fascículos cerebrales tienen la forma de *sheet-like* o lámina porque abarcan regiones más extensas de la corteza. Al atravesar la sustancia blanca, estas fibras no forman estructuras tubulares, como otros fascículos, sino que se disponen en forma de lámina. De manera general, se debe a que rodean estructuras cerebrales, adaptándose a la anatomía cerebral de manera más extendida y plana. Un ejemplo de estos fascículos son el cuerpo caloso y los fascículos cortos del cíngulo.

Esta limitación de no poder representar estas formas no es tan perjudicial para la evaluación realizada, ya que los algoritmos de *clustering* de fibras aplicados generan *clusters* tubulares. Sin embargo, sería muy útil extender el simulador en el futuro a otras formas de fascículos.

La aplicabilidad del simulador se demostró al generar tres conjuntos de datos simulados de fascículos de todo el cerebro con diferentes números de fascículos (y fibras) para evaluar el comportamiento de dos algoritmos: *QuickBundles* (QB) y *Fast Fiber Clustering* (FFClust), para diferentes umbrales de distancia. Se utilizaron cinco métricas para evaluar la calidad de los *clusters* obtenidos por los algoritmos. Con el análisis de estas métricas fue posible además evaluar

diferentes aspectos de los algoritmos, como el comportamiento ante cruzamientos y permutaciones de los datos de entrada, mostrando la complejidad de evaluar este tipo de algoritmos.

Capítulo 7

Base de datos simulados para la validación de un algoritmo de parcelación estructural

En este capítulo se describe la creación de una base de datos simulada de parcelas y sus conexiones. Esta base de datos fue generada a partir de una parcelación cortical aleatoria basada en distancia geodésica y la simulación de las conexiones entre las parcelas mediante una extensión del simulador de fascículos usando *spline*. La aplicabilidad de la base de datos se demostró en la validación de un algoritmo de parcelación cortical basado en conectividad estructural.

7.1. Métodos

En esta sección se describe la metodología utilizada en la obtención de la base de datos simulada. Para las simulaciones, se utilizaron los mallados y los datos de tractografía de 20 sujetos de la base de datos ARCHI [84]. Además la implementación del simulador basado en curvas *spline* se extendió para que

fuese capaz de manejar formas más complejas en los extremos como lo son la forma de las parcelas corticales.

La base de datos de conexiones simuladas se utilizó para la evaluación de un algoritmo de parcelación basado en conectividad estructural. Este algoritmo se validó comparando las parcelas generadas respecto a las de la parcelación basada en distancia geodésica que dio origen a las conexiones simuladas, usando métricas como el DICE y el Índice de Rand Ajustado. La metodología general se muestra en la Figura 7.1.1.

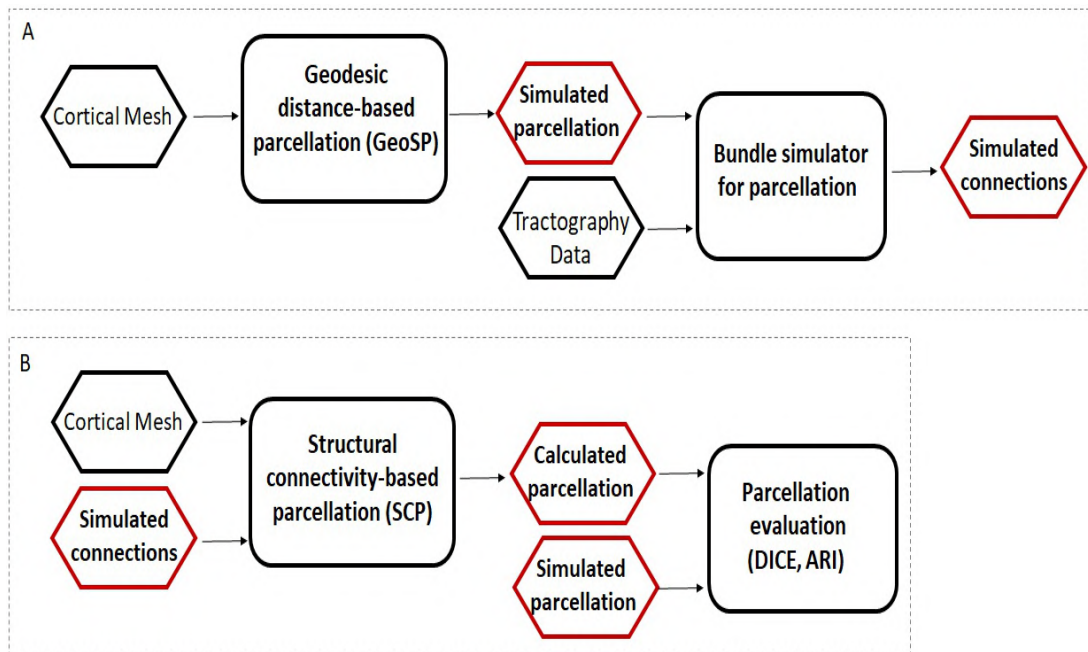


Figura 7.1.1: Esquema general de la secuencia de algoritmos aplicados en este capítulo para la generación de una base de datos simulada, utilizada en la validación del algoritmo de parcelación cortical. A: Generación de la base de datos simulada de parcelaciones y sus conexiones, B: Aplicación y validación de un algoritmo de parcelación cortical basado en conectividad estructural. Los cuadrados representan los algoritmos aplicados. Los polígonos representan los datos de entrada de cada algoritmo (polígonos en rojo representan salida de un algoritmo y entrada de otro)

7.1.1. Parcelación aleatoria basada en distancia geodésica

Para la obtención de las parcelas simuladas se utilizó una parcelación aleatoria basada en distancia geodésica (GeoSP [53]). GeoSP realiza una parcelación completa de la superficie cortical sin depender de un atlas anatómico predefinido, dividiendo la corteza en un número especificado de subparcelas usando distancia geodésica. Se utiliza la distancia geodésica ya que esta sigue el trayecto más corto sobre la superficie del mallado cerebral, respetando la topología de los giros (gyri) y surcos (sulci) de la corteza. Los giros son las elevaciones o pliegues de la superficie cortical, mientras que los surcos son las hendiduras que separan estos pliegues. La Figura 7.1.2 muestra un esquema general de la implementación del algoritmo.

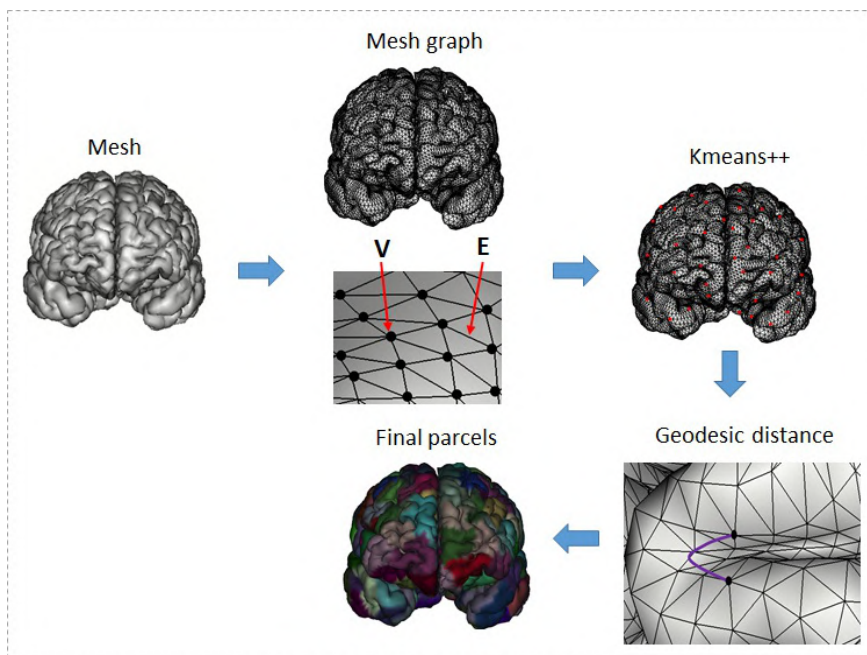


Figura 7.1.2: Diagrama del algoritmo de parcelación cortical basado en distancia geodésica. Representación en grafo del mallado de un sujeto: V representa los nodos del grafo y E las aristas. Para la inicialización de los centroides, se utiliza el algoritmo K-means++ para ubicar las semillas (rojo), y posteriormente, se emplea la distancia geodésica (magenta) para definir las parcelas finales.

Para realizar la parcelación completa de la corteza cerebral, GeoSP representa el mallado cortical como un grafo no dirigido $G(V, E)$. Donde cada nodo V

corresponde a un vértice del mallado cortical y cada arista E representa las conexiones entre estos vértices. A continuación, se aplica un algoritmo de *clustering* K-means basado en distancia geodésica, de manera paralela para cada hemisferio, siguiendo tres subpasos principales: (a) inicializar los centroides utilizando el algoritmo K-means++ para lograr una mejor distribución y una convergencia más rápida, (b) asignar los vértices al centroide más cercano en función de las distancias geodésicas calculadas mediante el algoritmo de Dijkstra [19, 8], y (c) recalcular los centroides, donde el nuevo centroide es el vértice más cercano a todos los demás vértices del *cluster*. Estos subpasos se repiten hasta alcanzar la convergencia, definida como un cambio en la posición del centroide inferior a 2 mm o un máximo de 20 iteraciones. La parcelación resultante asigna a cada vértice una etiqueta correspondiente a su parcela respectiva.

Finalmente, se generó una parcelación única a partir de un sujeto de la base de datos ARCHI [84] con 150 parcelas por hemisferio. Dado que la ubicación de los triángulos en los mallados corticales de los sujetos corresponden a las mismas áreas anatómicas, esta parcelación se aplicó a los restantes 19 mallados de los sujetos. De este modo, en la base de datos simulada se tendrá en cuenta únicamente la variabilidad natural de la morfología de los sujetos.

7.1.2. Simulación de las conexiones entre parcelas

Para generar la base de datos de conexiones simuladas se parte de la simulación de los fascículos que conectan un par de parcelas, utilizando una extensión del simulador basado en curvas *spline* (ver Sección 6.1.1). El simulador mantiene el formato tubular en las regiones intermedias pero se modifica la forma en la que se simulan los extremos para que sean correspondientes con la forma que tienen las parcelas. Los parámetros de entrada del simulador incluyen una fibra centroide, obtenida mediante un algoritmo de intersección entre los datos tractografía y el

mallado cortical de cada sujeto de la base de datos, y los puntos de las regiones externas, que se derivan de cada par de parcelas determinadas por el algoritmo de parcelación basado en distancia geodésica. La Figura 7.1.3 muestra la metodología general de la generación de la base de datos.

7.1.2.1. Preprocesamiento de los datos de entrada

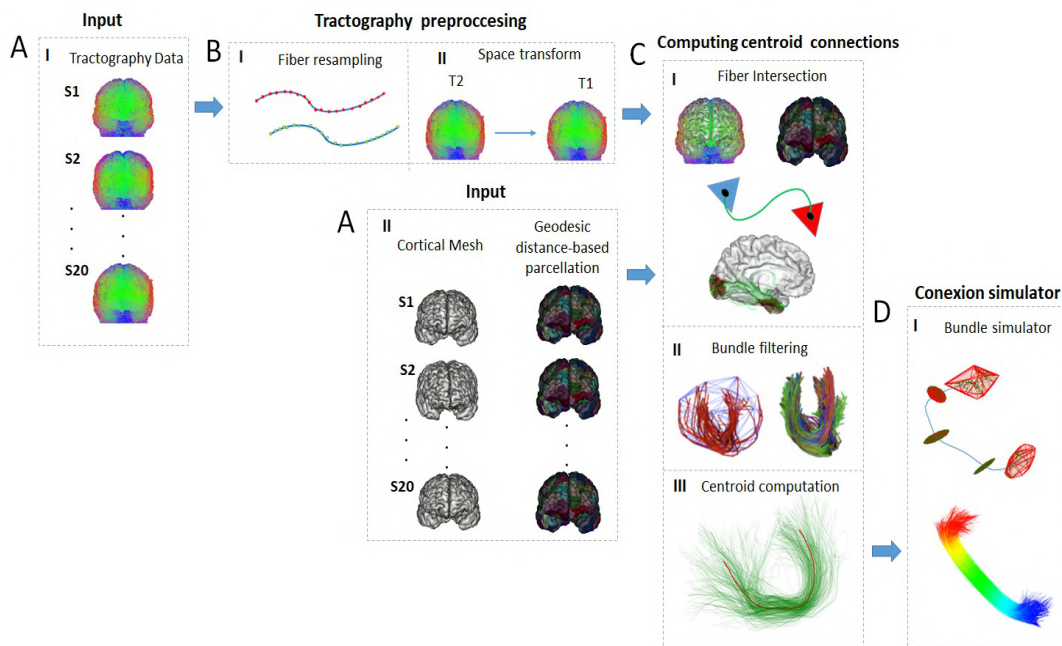


Figura 7.1.3: Esquema general para la simulación de las conexiones. A) Datos de entrada, B) Preprocesamiento de los datos, C) Obtención de los centroides, D) Simulador de fascículos.

A los sujetos de la base de datos se les realiza un preprocesamiento como se muestra en la Figura 7.1.3 B.-I-II que comienza con el remuestreo de todas las fibras de los datos de tractografía a 21 puntos. Esto es necesario porque las fibras pueden tener diferente cantidad de puntos, y el remuestreo facilita los análisis posteriores [96, 94, 73]. El siguiente paso es una transformación de espacio: los datos de tractografía, inicialmente en el espacio T2, se transforman al espacio del mallado de los sujetos, correspondiente al espacio T1.

7.1.2.2. Intersección de la tractografía con el mallado cortical

Para determinar los fascículos que intersectan un par de parcelas se utilizó un algoritmo de intersección de la tractografía con el mallado cortical [76] (ver Figura 7.1.3.C-I). Este algoritmo calcula la intersección entre las fibras y el mallado cortical usando el método de intersección rayo-triángulo de Möller-Trumbore [62], el cual determina si un rayo atraviesa un triángulo. En este caso, se define un rayo por cada extremo de las fibras, mientras que el triángulo está formado por los tres vértices del mallado cortical.

El algoritmo evalúa en cada iteración si los extremos de cada fibra intersectan algún triángulo de la superficie cortical, además considera la división del espacio 3D en celdas cúbicas, lo que acelera la búsqueda de intersecciones entre una fibra y un triángulo de la malla. Además, el algoritmo proyecta las fibras un punto hacia atrás (por si se extienden más allá de la corteza) y dos puntos hacia adelante (por si no alcanzan la corteza), asegurando así que no se pierdan intersecciones en los bordes de las celdas.

El algoritmo genera un archivo con información detallada sobre las fibras que intersectan triángulos del mallado, que incluye el índice de cada fibra, el hemisferio en el que se produce la intersección en cada extremo, los puntos exactos de intersección inicial y final, y los triángulos correspondientes en la malla cortical.

7.1.2.3. Obtención de los centroides que conectan un par de parcelas

Como resultado del algoritmo de intersección se obtiene un perfil de conectividad indicando cuáles fascículos de fibras conectan un par de parcelas. Estos fascículos por limitaciones propias de la reconstrucción de las fibras durante el algoritmo de tractografía son fascículos que pueden presentar fibras ruidosas. Para filtrar estas fibras se utiliza como preprocesamiento un algoritmo de filtrado por envolvente

convexa [57].

Este filtro de fascículos utiliza la envolvente convexa para generar el menor cascarón convexo que encierra al fascículo como se muestra en la Figura 7.1.3.C-II. Para aplicar la envolvente convexa el fascículo se representa como una nube de puntos. A cada fibra del fascículo con al menos un punto en la envolvente, se define un Grado de Anormalidad (*Degree of Abnormality, DA*). Este DA se calcula midiendo la distancia Euclidiana promedio de cada punto de la fibra a sus K vecinos más cercanos de la nube de puntos. Las fibras con un DA superior a una desviación estándar del promedio se eliminan iterativamente hasta descartar un porcentaje de fibras. En este trabajo se utilizaron los parámetros establecidos por [Mendoza et al.\(2024\)](#) de porcentaje de eliminación del 20% y un número de vecinos de 80.

Posterior al filtrado de los fascículos se realizó el alineamiento de todas las fibras del fascículo con respecto a una de referencia. La alineación de las fibras del fascículo es necesaria para evitar que se almacenen en orden directo o inverso. Para alinear cada fibra, se compara su orientación con una fibra de referencia, reorientando la fibra si la distancia entre extremos opuestos es mayor. La fibra de referencia se selecciona eligiendo la de menor distancia a las restantes fibras. Finalmente, se calcula el centroide del fascículo como la media de sus fibras.

También es necesario determinar cuál es la parcela de inicio del centroide y cuál la de término. Para este propósito, se utiliza la dirección de la fibra de referencia, dado que todas las fibras han sido alineadas con respecto a ella. La información sobre los triángulos iniciales y finales que intersectan las fibras, proporcionada por el algoritmo de intersección, se analiza para determinar a qué parcela pertenece cada triángulo. Así, la dirección del centroide corresponde a la dirección de la fibra de referencia.

El centroide obtenido del fascículo puede intersectar un triángulo en cualquier

posición dentro de la parcela. Para trasladar el centroide hacia el centro de la parcela con el objetivo de obtener un fascículo con una forma más realista y homogénea, y a la vez preservar su forma, se realizan una serie de transformaciones geométricas basadas en una combinación de rotación y escalado.

Primero se calcula una matriz de rotación basada en la fórmula de Rodrigues [79] (ver Eq. 6.1.1) utilizando los vectores de dirección entre los puntos extremos del centroide original y los nuevos puntos de inicio y final. Esta rotación alinea el centroide con la nueva orientación espacial de los puntos centrales de la parcela. Finalmente, se aplica un escalado uniforme si es necesario, ajustando el tamaño del centroide sin distorsionar su forma, para que los puntos finales coincidan con las distancias requeridas en la nueva ubicación. La Figura 7.1.3.C-III muestra un ejemplo de la traslación de los centroides.

7.1.2.4. Simulador de fascículos

Para simular un fascículo se utilizó una extensión del simulador basado en curvas *spline* presentado en la Sección 6.1.1. Se mantiene para la simulación la división del fascículo en cinco regiones transversales: las dos regiones extremas, la región central y dos regiones intermedias. La metodología seguida se basa en representar en las regiones externas las formas reales de las parcelas, mientras que las regiones intermedias y centrales conserven una estructura circular. Una representación del modelo se muestra en la Figura 7.1.3.D-I. Los parámetros claves para el proceso de simulación del fascículo en este modelo fueron el centroide, que proporciona una descripción aproximada de la trayectoria del fascículo, los vértices de las dos parcelas que conecta el centroide y el número de fibras a simular.

La forma de la parcela está definida por los triángulos que la componen. Basado en esta idea, se decidió utilizar los vértices de los triángulos de la parcela como puntos extremos de las regiones del fascículo a simular. Para las regiones intermedias y la

central, se mantuvo el enfoque de emplear círculos centrados en sus respectivos puntos centroides (ver Sección 6.1.1.2). En las regiones intermedias, los radios se obtuvieron promediando los radios de la región externa más cercana, mientras que para la región central se usó un radio que permaneciera dentro del rango de los radios de las regiones intermedias, sin exceder ninguno de ellos.

Es importante destacar que de manera general diferentes parcelas tienen diferentes número de vértices. Por tanto, el número de fibras a simular tiene que ser mayor o igual al número de vértices de la parcela mayor. Para la parcela de menor número de vértices, o si el número de fibras a simular supera al de la parcela mayor, para ambos casos se realiza una interpolación lineal con los puntos que componen los vértices para que sea igual el número de fibras a simular.

Para cada región intermedia y central, el algoritmo genera puntos siguiendo una distribución aleatoria uniforme. Esta elección se fundamenta en el análisis realizado sobre varios fascículos del atlas de DWM [39], donde los puntos de estas regiones mostraron, de manera visual, un patrón consistente con una distribución uniforme. Además, para confirmar esta observación, se calculó el coeficiente de dispersión [43], una métrica que evalúa la uniformidad espacial de los datos. Los resultados obtenidos, con valores cercanos a cero, indicaron que los puntos tienden a distribuirse de forma uniforme.

Para la representación de las fibras cerebrales se utilizan curvas *spline*. Con el simulador de *spline* que se presentó en el Capítulo 6 solo se simulaban fibras largas mayores a 50 mm. En esta aproximación extendida del simulador se incorpora la simulación de fibras cortas ya que existen parcelas muy cercanas que están unidas por fibras cortas. Considerando esto y que de manera general la forma de las fibras cortas tiende a ser en forma de U o con menos curvaturas que las presentes en las fibras largas. Entonces, para la representación de las fibras cortas se utilizan *spline* de segundo grado basadas en solo 3 puntos de control obtenidos de las

regiones externas y central y para fibras largas se usan *spline* de cuarto orden basadas en 5 puntos de control obtenidos de las regiones externas, intermedias y central. Un ejemplo que ilustra la forma de las fibras que conectan un par de parcelas se presenta en la Figura 7.1.4. En el primer panel, se observa una fibra corta con una forma de U que conecta dos parcelas muy cercanas. Por otro lado, en el segundo panel, se muestra una fibra con una forma más compleja, ya que conecta dos parcelas distantes entre sí.

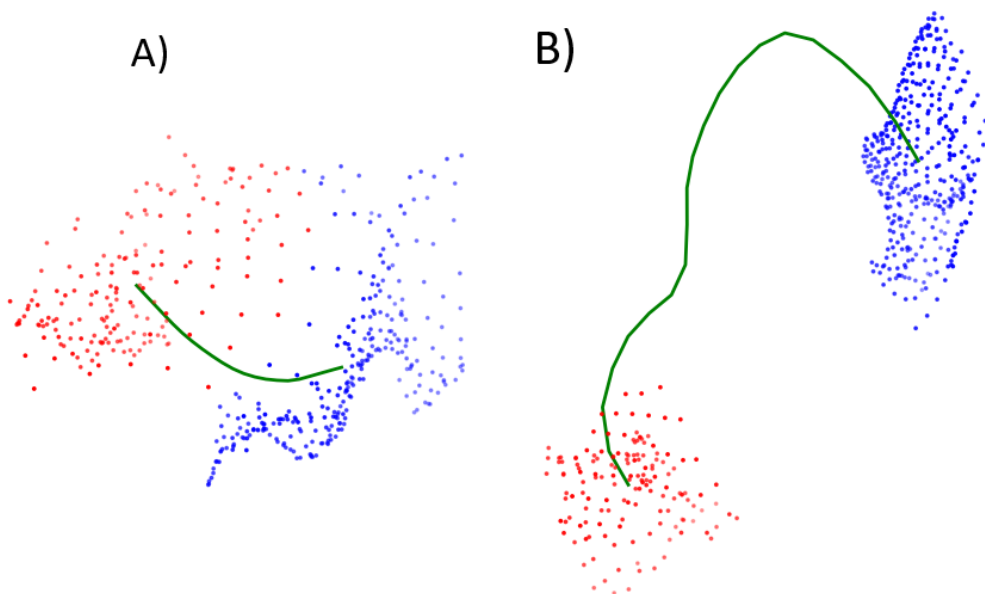


Figura 7.1.4: Ejemplos de centroides que conectan un par de parcelas corticales. A: fibra corta que conecta dos parcelas cercanas, B: fibra larga que conecta dos parcelas lejanas.

Para introducir dispersión en los extremos de la curva, se puede añadir opcionalmente ruido gaussiano a los puntos terminales de la *spline*. Finalmente, para generar la base de datos de conexiones simuladas, se utilizan los centroides obtenidos de la intersección y los vértices de las parcelas que conectan los centroides. Con esta información, se simulan 20 datos de tractografía de cerebro completo utilizando el simulador de fascículos.

7.1.3. Validación de un algoritmo de parcelación basado en conectividad estructural

La base de datos simulada se emplea para evaluar el desempeño de un algoritmo de parcelación cortical basado en conectividad estructural propuesto en el estado del arte [94]. Este algoritmo toma como entrada los datos de tractografía simulados y los mallados corticales de los distintos sujetos. La parcelación resultante se compara con la generada por el algoritmo GeoSP, que fue utilizada previamente para crear las conexiones simuladas. Para evaluar el desempeño del algoritmo de parcelación, se utilizan el coeficiente de DICE y el Índice de Rand Ajustado. La Figura 7.1.5 muestra los pasos generales para la validación del algoritmo.

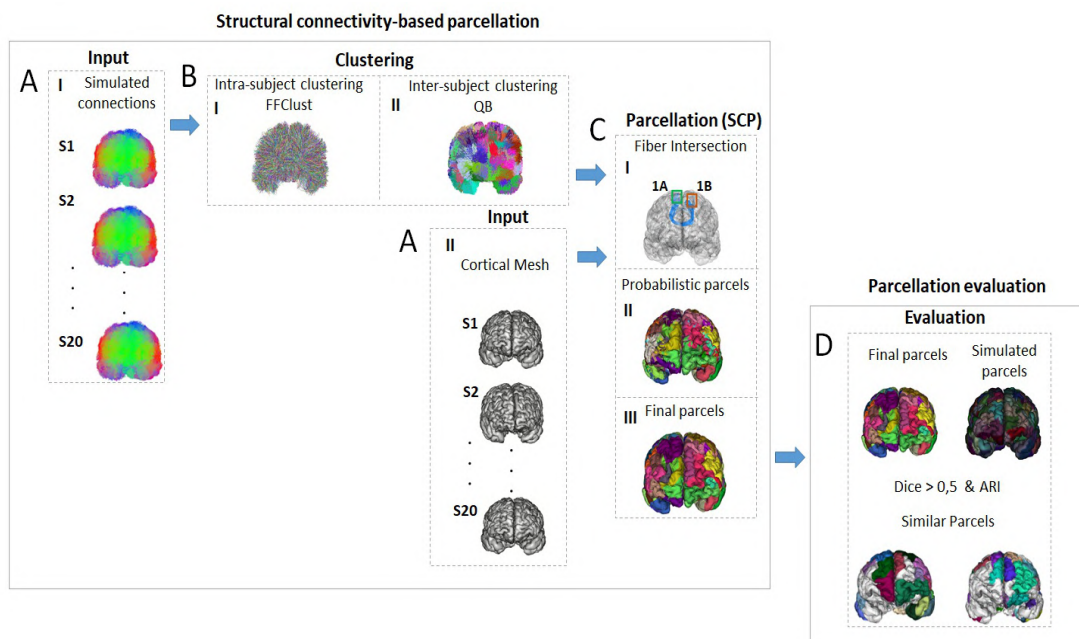


Figura 7.1.5: Diagrama general del proceso de validación de un algoritmo de parcelación cortical basado en conectividad estructural. A-C) Algoritmo de parcelación cortical. D) Comparación entre las parcelaciones simuladas y las generadas por el algoritmo basado en conectividad estructural.

El algoritmo de parcelación que se evalúa en esta tesis es el propuesto por Vergara et al. (2021). Este algoritmo parte con un *clustering* de todas las fibras del sujeto usando FFClust para obtener los centroides representativos por sujeto. Luego,

se aplica un *clustering* de fibras inter-sujeto mediante QB para identificar los fascículos representativos de la población. Para ajustar los parámetros de FFClust y QB, se analizó qué combinación de parámetros lograba una mejor recuperación del *ground truth* (conexiones simulada). La metodología para la evaluación de los algoritmos de *clustering* fue descrita en la Sección 6.1.3. Los umbrales de distancia evaluados para FFClust fueron $dR_{max} = d_{Mmax} = (6, 12, 18, 20, 25)$ mm, y para determinar el tamaño de los *clusters* de puntos se utilizó el método del codo como fue propuesto por Vázquez et al. (2019). En cuanto a QB, se emplearon umbrales de distancia $\theta_{QB} = (10, 12, 15)$ mm. La combinación que alcanzó un mayor *Accuracy* fue $dR_{max} = d_{Mmax} = 20$ mm y $\theta_{QB} = 20$ mm, obteniendo un valor de 0.50. Posteriormente, se filtraron los *clusters* inter-sujeto obtenidos según un umbral de representatividad de sujetos thr_{rp} para conservar solo los *clusters* que fueran representativos de la población.

Para obtener las parcelas preliminares, los *clusters* representativos se intersectan con el mallado cortical, y cada extremo define una parcela preliminar. Idealmente, cada *cluster* debería terminar en una región cortical específica. Sin embargo, muchos triángulos del mallado son intersectados múltiples veces por fibras de diferentes *clusters* o incluso por fibras del mismo *cluster*. Para abordar esto, se generan mapas de densidad cortical que muestran la frecuencia de aparición de cada parcela preliminar en los triángulos, y se filtran las parcelas pequeñas para simplificar el proceso.

En la siguiente etapa, a cada triángulo se le asigna la etiqueta de la parcela con la mayor probabilidad en función de un umbral de probabilidad thr_{dc} . Además, se resuelve el traslape entre subparcelas mediante un grafo no dirigido. Las parcelas se consideran significativamente solapadas cuando el traslape entre sus centros de densidad supera un umbral de intersección thr_{it} . Aquellas parcelas con una intersección significativa en sus centros de densidad se fusionan, obteniendo de

esta manera las parcelas probabilísticas definitivas.

Para obtener las parcelas finales, se aplica un proceso de post-procesamiento sobre las parcelas probabilísticas, debido a que estas suelen presentar grupos de triángulos aislados y vacíos en diversas regiones. En primer lugar, se calcula la componente conexa más grande de cada parcela, que se toma como representativa. A continuación, se realiza una apertura morfológica que permite corregir las irregularidades y generar parcelas finales más robustas. Esta apertura consiste en una erosión seguida de una dilatación: la erosión reduce los triángulos en los bordes de una región, mientras que la dilatación, como operación inversa, expande las regiones en sus bordes [87, 94, 76]. En este trabajo se utilizó un factor de erosión de 1 y la dilatación con un factor de 4 como fue propuesto en Vergara et al. (2021).

Se realizaron múltiples pruebas para evaluar el comportamiento del algoritmo, variando los parámetros utilizados en la obtención de la parcelación. En total, se llevaron a cabo seis pruebas con el objetivo de analizar la sensibilidad del algoritmo ante la variación de dichos parámetros. Esto es importante pues en el trabajo propuesto por Vergara et al. (2021), los valores de los parámetros fueron ajustados heurísticamente. Los umbrales fueron modificados de la siguiente manera: $thr_{rp} = (10, 12)$, $thr_{dc} = (0.2, 0.4)$ y $thr_{it} = (0.1, 0.3)$.

Para evaluar el rendimiento del algoritmo en las diferentes pruebas, se realizó una comparación entre las parcelas calculadas y las simuladas utilizando dos métricas ampliamente reconocidas en el estado del arte: el coeficiente DICE y el Índice de Rand Ajustado (ARI). DICE se usa para medir el solapamiento entre las parcelas de dos parcelaciones. Dadas dos parcelas, X e Y , este coeficiente se calcula mediante la siguiente fórmula:

$$\text{Dice} = \frac{2|X \cap Y|}{|X| + |Y|} \quad (7.1.1)$$

donde $|\cdot|$ denota el número de vértices en cada parcela. Sin embargo, una limitación de esta métrica es que puede sesgar los resultados al comparar parcelas de tamaño muy desigual, favoreciendo ciertos tipos de parcelaciones [2]. Un coeficiente de Dice de 1 implica una coincidencia perfecta (parcelaciones idénticas). Para calcular el DICE, se compararon las parcelaciones generadas por el algoritmo basado en conectividad estructural con aquellas obtenidas mediante el algoritmo basado en distancia geodésica, identificando las parcelas cuyo coeficiente DICE fuera superior a 0.5.

ARI se basa en contar el número de elementos sobre los cuales dos parcelaciones están de acuerdo o en desacuerdo [2]. Clasifica los pares de triángulos $\binom{N}{2}$ en uno de los cuatro conjuntos $(N_{11}, N_{00}, N_{01}, N_{10})$, basándose en su etiquetado en cada parcelación. Para las parcelaciones U y V , N_{11} corresponde al número de pares que se asignan a la misma parcela en ambas, N_{00} corresponde al número de pares que se asignan a diferentes parcelas, N_{01} corresponde al número de pares que se asignan a la misma parcela en U , pero a diferentes parcelas en V , y N_{10} corresponde al número de pares que se asignan a la misma parcela en V , pero a diferentes parcelas en U . Intuitivamente, N_{00} y N_{11} representan el acuerdo entre las parcelaciones, mientras que N_{01} y N_{10} indican su desacuerdo. Después de contar el número de pares, el ARI para las parcelaciones U y V se calcula como se muestra en la ecuación . 7.1.2. Un ARI de 1 indica una correspondencia perfecta entre las parcelaciones, mientras que un valor de 0 implica que las parcelaciones no coinciden en ninguna de las etiquetas.

$$\text{ARI}(U, V) = \frac{2(N_{00}N_{11} - N_{01}N_{10})}{(N_{00} + N_{01})(N_{01} + N_{11}) + (N_{00} + N_{10})(N_{10} + N_{11})} \quad (7.1.2)$$

7.2. Resultados

7.2.1. Generación de la base de datos de parcelas simuladas

Se generó una parcelación basada en distancia geodésica utilizando la malla de un sujeto de la base de datos ARCHI. Dado que los triángulos de las mallas de los diferentes sujetos presentan correspondencia anatómica, se calcula una parcelación para el sujeto inicial y, posteriormente, se aplica esta misma parcelación a los restantes sujetos. Si bien los vértices son los mismos, existen variaciones en la morfología de las parcelas entre sujetos. Este enfoque garantiza que en la base de datos de parcelaciones se capture la variabilidad inherente a los diferentes mallados de los sujetos.

La Figura 7.2.1 ilustra la parcelación aplicada a los mallados de los tres primeros sujetos de la base de datos. Los resultados muestran parcelas homogéneas en ambos hemisferios. En cuanto a la variabilidad observada entre los sujetos, al mantener una misma parcelación calculada a partir de un sujeto y aplicarla a diferentes mallados, se puede apreciar que la morfología de los sujetos varía ligeramente. Esta variación provoca que las parcelas se desplacen sutilmente, logrando capturar la variabilidad anatómica deseada, haciéndola factible para análisis multisujeto.

7.2.2. Generación de la base de datos de conexiones simuladas

Para la generación de la base de datos de conexiones simuladas, se utilizaron como parámetros de entrada los centroides obtenidos de los 20 conjuntos de datos de tractografía de la base de datos ARCHI, junto con los vértices de las parcelas extraídos de las parcelaciones corticales simuladas. La Figura 7.2.2 muestra un ejemplo de un fascículo simulado conectando dos parcelas. Al utilizar los vértices

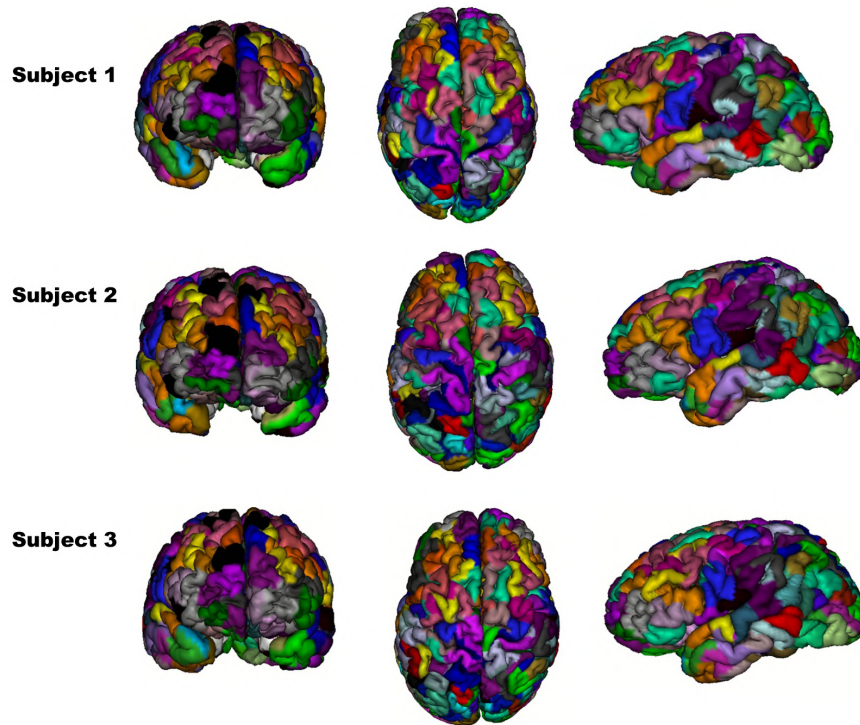


Figura 7.2.1: Parcelación GeoSP visualizada en tres sujetos de la base de datos ARCHI. A la izquierda se encuentra la vista posterior, en el centro la vista dorsal y a la derecha la vista lateral.

como puntos de las regiones externas, se garantiza que el fascículo refleje con precisión la forma de la parcela en ambos extremos y logre cubrir la parcela completamente.

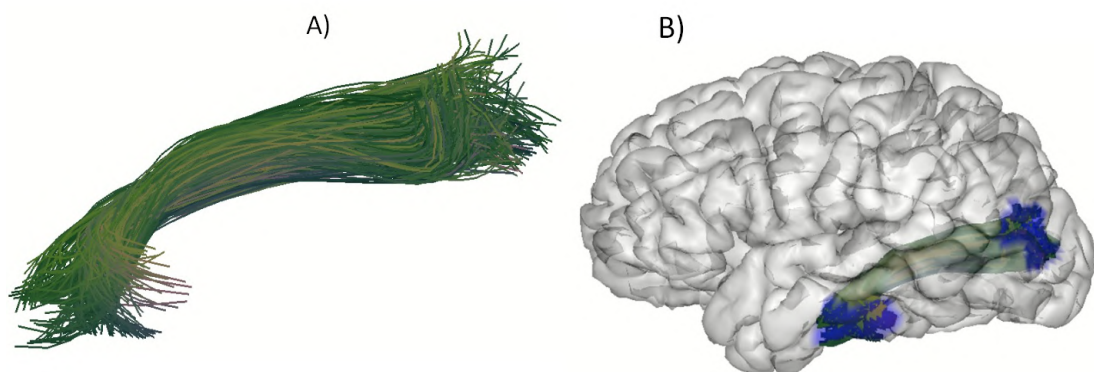


Figura 7.2.2: Ejemplo de un fascículo simulado a partir del simulador. A) muestra el fascículo simulado. B) Muestra el fascículo simulado (verde) y las parcelas que conecta (azul).

La Figura 7.2.3 presenta las conexiones simuladas de los tres primeros sujetos, con 916,057 fibras (1,995 centroides), 805,965 fibras (1,757 centroides) y 908,650 fibras (1,966 centroides), respectivamente. En estos resultados puede apreciarse que la forma de las conexiones simuladas varían para los diferentes sujetos, lo que está en correspondencia con la variación morfológica presente en los mallados de los sujetos y que se había visto en los resultados de la parcelación cortical. Es esperable ver esta variación presente en las conexiones simuladas ya que la información de la parcelación cortical sirve como base para la obtención de estas.

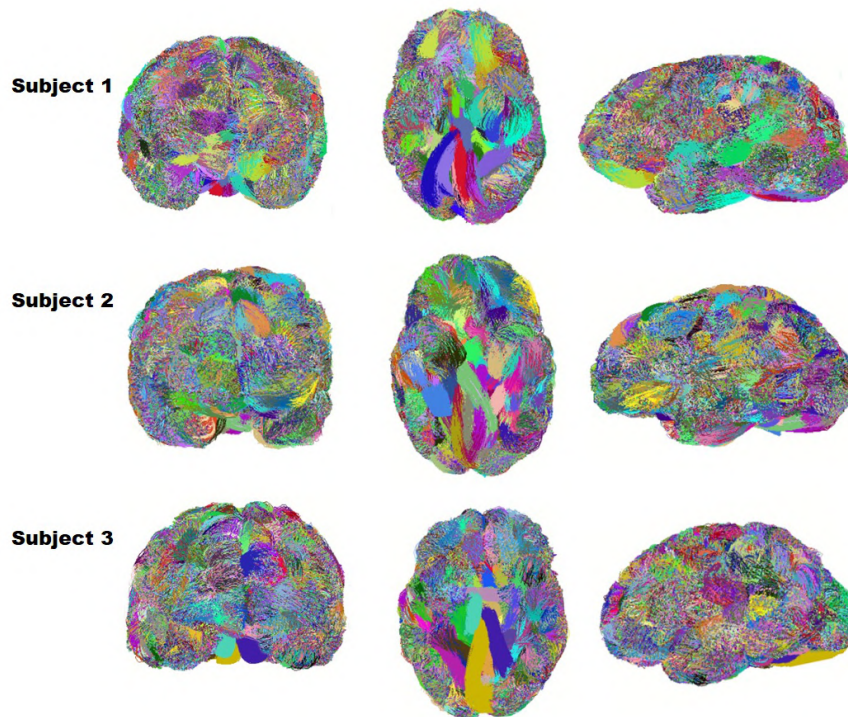


Figura 7.2.3: Conexiones simuladas de tres sujetos de la base de datos. Colores aleatorios representan los diferentes fascículos que componen las conexiones simuladas.

7.2.3. Validación del algoritmo de parcelación cortical usando la base de datos simulada

La base de datos de parcelación, es decir las parcelas más las conexiones entre estas, se utilizó para la validación de un algoritmo de parcelación basado en

conectividad estructural. La parcelación estructural se comparó con la parcelación basada en distancia geodésica que se utilizó como base para generar las conexiones simuladas. Para el proceso de validación se evaluó el traslape y la correspondencia de las parcelas usando el coeficiente de DICE y el índice de Rand Ajustado.

Este algoritmo basado en conectividad estructural es dependiente de varios parámetros, como los umbrales de distancia de los algoritmos de *clustering* de fibras, los umbrales de representatividad de los sujetos (thr_{rp}), de probabilidad e intersección (thr_{dc}, thr_{it}). En el caso de los parámetros correspondientes a los algoritmos de *clustering*, como se mencionó en la Sección 7.1.3, fueron determinados y fijados mediante la evaluación de los algoritmos de *clustering* en cuanto a la recuperación del *ground truth*. Finalmente se realizaron seis pruebas para evaluar el comportamiento del algoritmo de parcelación cortical variando los parámetros: thr_{rp} , thr_{dc} y thr_{it} .

La Tabla 7.2.1 presenta los resultados obtenidos para las diferentes combinaciones de parámetros, mostrando el número total de parcelas por hemisferio, el número de parcelas similares por hemisferio considerando un DICE >0.5 , así como los resultados del Índice de Rand Ajustado (ARI). En azul se destaca la mejor combinación de parámetros, la cual corresponde a $thr_{dc} = 0.4$, $thr_{it} = 0.3$ y $thr_{rp} = 10$, resultando en 178 parcelas en el hemisferio izquierdo (118 similares) y 190 en el hemisferio derecho (120 similares), con un ARI de 0.61. Los resultados de similitud medidos por el DICE indican que el algoritmo es capaz de identificar alrededor del 80% de las parcelas del *ground truth*, que en este caso se refiere a la parcelación obtenida mediante distancia geodésica.

Como resultado de incrementar el umbral de representatividad a 12, se obtiene un mayor número de parcelas similares, pero también un mayor número de parcelas totales, manteniendo el mismo ARI que la combinación anterior. Esto se debe a que el umbral de representatividad de sujetos afecta el número de *clusters*

inter-sujeto a ser considerados para la intersección, información utilizada para la construcción de las parcelas preliminares. Al aumentar el umbral, se considera un menor número de fascículos. Las conexiones descartadas influyen en la definición de las parcelas finales, ya que aparecen un mayor número de parcelas pequeñas que, al considerar más conexiones, solían formar una única parcela. Este efecto conlleva a un incremento en el número de parcelas finales.

Parámetros			Resultados					
thr_{rp}	thr_{dc}	thr_{it}	PF LH	PF RH	PS LH	PS RH	ARI LH	ARI RH
10	0.4	0.1	192	205	118	119	0.52	0.51
10	0.2	0.1	157	167	108	111	0.51	0.51
10	0.4	0.2	176	188	116	119	0.55	0.55
10	0.4	0.3	178	190	118	120	0.60	0.61
10	0.4	0.4	178	191	118	120	0.60	0.61
12	0.4	0.3	216	234	126	125	0.60	0.61

Tabla 7.2.1: Resultados de la parcelación basada en conectividad estructural al variar los parámetros del algoritmo: umbral de representatividad de sujetos thr_{rp} , umbral de probabilidad thr_{dc} , umbral de intersección thr_{it} . Parcelas finales (PF) por hemisferio: izquierdo (LH) y derecho (RH). Adicionalmente, se indica la cantidad de parcelas similares (PS) por hemisferio (DICE >0.5) y el valor del Índice de Rand Ajustado. La mejor combinación de parámetros se destaca en azul.

La Figura 7.2.4 presenta los resultados de la aplicación de los algoritmos de *clustering*. En la Figura 7.2.4.A se observan los centroides obtenidos mediante FFClust, mientras que la Figura 7.2.4.B ilustra los *clusters* inter-sujeto resultantes de agrupar los centroides de FFClust utilizando el algoritmo QB y aplicando un umbral de representatividad de sujetos para definir los *clusters* finales. La Figura 7.2.5 muestra los resultados de las distintas etapas del algoritmo de parcelación, estas son las parcelas preliminares obtenidas de la intersección, las parcelas probabilísticas resultantes de los mapas de densidad y las parcelas finales después de aplicar las operaciones morfológicas.

En la parcelación final se observan algunos espacios en blanco que corresponden a triángulos no etiquetados en ninguna parcela. Esto ocurre debido al filtrado de

parcelas pequeñas realizado después de definir las parcelas preliminares, lo que provoca la pérdida de información de algunas de estas. Para evitar este problema, es necesario implementar, antes del proceso de filtrado, un algoritmo que permita fusionar estas parcelas pequeñas con las parcelas grandes más cercanas, asegurando que la mayoría de las áreas queden etiquetadas.

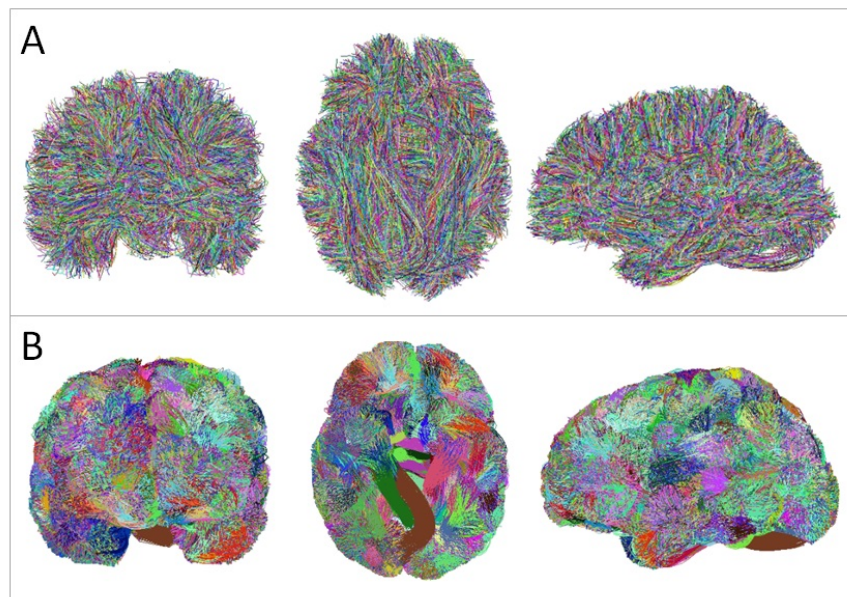


Figura 7.2.4: Resultados de la aplicación de los algoritmos de *clustering* para un sujeto de la base de datos: A) Centroides de los *clusters* intra-sujeto agrupados por FFClust, B) Centroides intrasujetos agrupados por QuickBundles (cada color representa un *cluster* inter-sujeto).

Otro aspecto relevante son las parcelas similares obtenidas mediante el coeficiente DICE. La Figura 7.2.6 muestra, en la parte superior, las parcelas simuladas que son similares al algoritmo de parcelación basado en conectividad, mientras que en la parte inferior se presentan las correspondientes a dicho algoritmo. Del análisis de la figura se observa que espacialmente y en tamaño existe una buena correspondencia de las parcelas similares en ambos algoritmos.

Un análisis adicional es el comportamiento de aquellas parcelas que difieren entre ambos algoritmos. En la Figura 7.2.7 se muestran las parcelas diferentes del *ground truth* con respecto a las que encuentra el algoritmo. La figura muestra la presencia

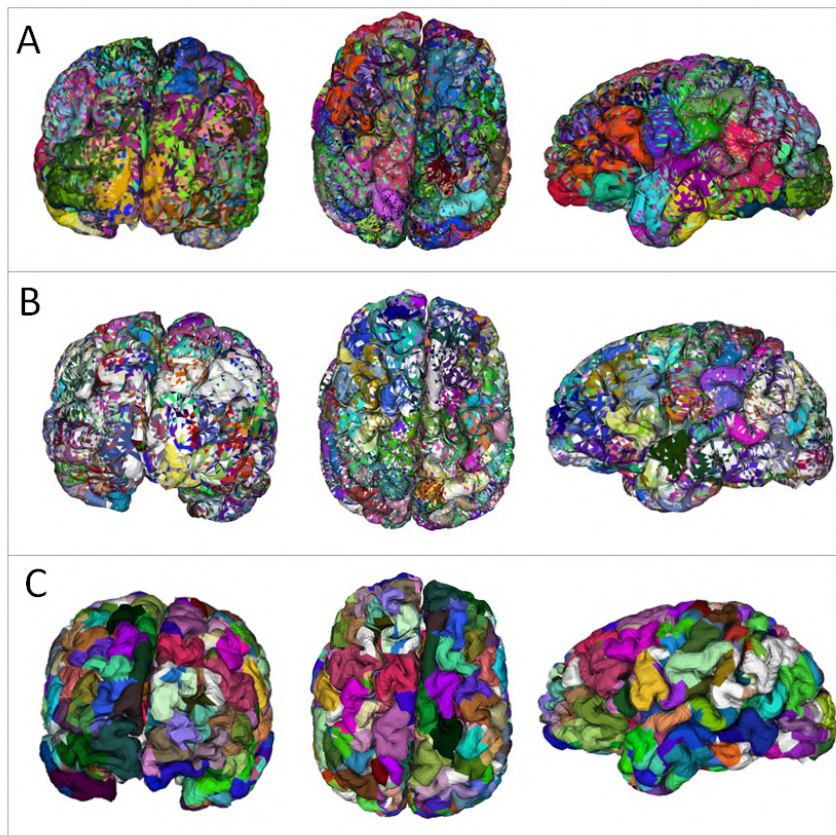


Figura 7.2.5: Resultados de las diferentes etapas del algoritmo de parcelación cortical basado en conectividad estructural. A) Parcelas Preliminares, B) Parcelas Probabilísticas, C) Parcelas finales.

de muchas parcelas pequeñas generadas por el algoritmo basado en conectividad, que podrían filtrarse o fusionarse con otras.

Se deriva de los análisis realizados la importancia de la base de datos simulada como *ground truth* para evaluar el comportamiento de los algoritmos de parcelación cortical basados en conectividad estructural, ya que no solo puede funcionar para validación sino también para incorporar mejoras y hacer los algoritmos más robustos.

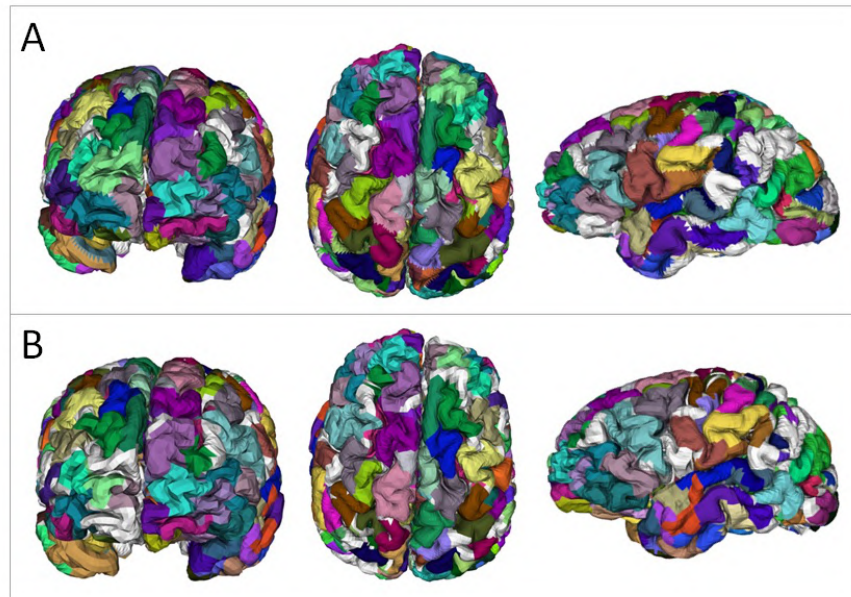


Figura 7.2.6: Resultados de las parcelas similares entre ambas parcelaciones para un DICE >0.5 . A) Parcelas simuladas (*ground truth*), B) Parcelas generadas por el algoritmo basado en conectividad estructural.

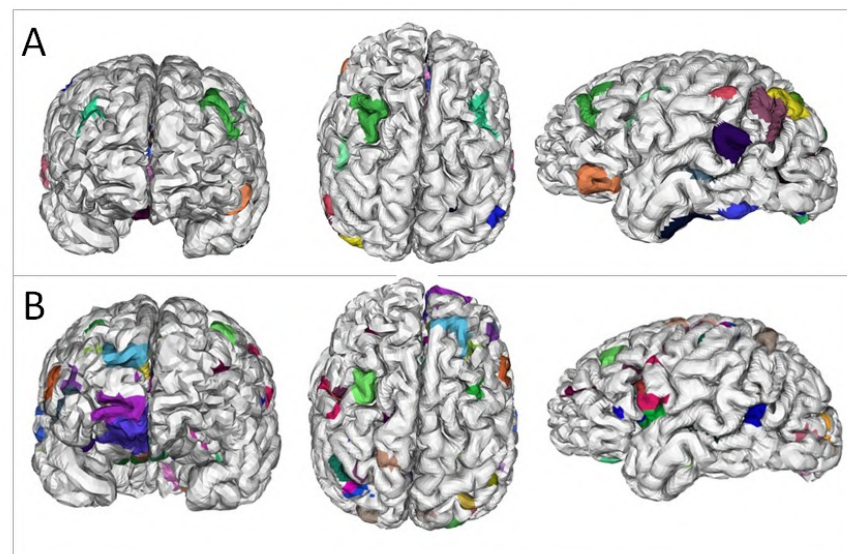


Figura 7.2.7: Resultados de las parcelas diferentes entre ambas parcelaciones. A) Parcelas generadas por GeoSP (*ground truth*), B) Parcelas generadas por el algoritmo basado en conectividad estructural.

7.3. Discusión

En este trabajo se generó una base de datos de parcelaciones corticales y sus conexiones a partir de una parcelación basada en distancia geodésica y un simulador de fascículos. El simulador de fascículos utiliza los centroides obtenidos de fibras de sujetos y mallados corticales con correspondencia anatómica entre sujetos, permitiendo generar datos para múltiples individuos. El simulador se adapta a las características de cada región cortical convirtiéndolo en una herramienta flexible y útil para evaluar y mejorar algoritmos de parcelación cortical basados en la conectividad estructural del cerebro. La base de datos simulada se utilizó para evaluar el comportamiento de un algoritmo de parcelación cortical basado en conectividad estructural ante diferentes configuraciones de sus parámetros de entrada.

Sin embargo, el simulador presenta algunas limitaciones como que parte de una parcelación inicial específica basada en el mallado de un único sujeto. Además, aunque el simulador genera fibras basadas en centroides, se podrían incorporar simulaciones de fibras más complejas, con parámetros adicionales como traslapes y ruido. Superar estas limitaciones en trabajos futuros permitirá desarrollar simulaciones aún más realistas y útiles para la validación de algoritmos.

También como trabajo futuro, sería interesante incluir zonas subcorticales y adoptar parcelaciones más realistas basadas en atlas anatómicos o en las circunvoluciones corticales como base para la generación de las conexiones simuladas, además de simular variaciones anatómicas entre sujetos.

Capítulo 8

Discusión y Conclusiones

Los métodos de análisis de datos de tractografía como los métodos de *clustering* de fibras y de parcelación de la corteza son de gran importancia en la descripción del conectoma. Los algoritmos de *clustering* de fibras tienen un papel importante por su capacidad para identificar automáticamente fibras similares mientras que los de parcelación de la corteza cerebral permiten subdividir la corteza cerebral en un conjunto de regiones que representen la conectividad anatómica y funcional del cerebro. Sin embargo, a pesar de su utilidad, estos algoritmos son difíciles de evaluar, comparar y mejorar debido a la ausencia de un *ground truth*.

Una alternativa eficaz para evaluar el comportamiento de estos algoritmos es contar con datos de fibras simulados. Sin embargo, los fascículos de fibras cerebrales presentan, en general, una forma compleja debido a la disposición intrincada de las fibras que los componen, cuyos extremos suelen ser dispersos y desordenados. Esto convierte la tarea de lograr una representación realista en un desafío significativo.

La simulación de fascículos de fibras se llevó a cabo mediante dos enfoques principales. El primero utiliza curvas exponenciales en los extremos para representar las fibras. Este enfoque es una aproximación inicial para evaluar

la factibilidad de utilizar simuladores de fibras. En contraste, el segundo enfoque, al utilizar curvas *spline*, basadas en polinomios, permite una representación más precisa de las fibras cerebrales. Esta técnica no solo es útil para la evaluación de algoritmos de *clustering*, sino también para algoritmos más avanzados, como los utilizados en la parcelación cortical basada en la conectividad estructural.

Para validar la precisión de ambas aproximaciones del simulador, se replicaron fascículos correspondientes a un atlas de fascículos de DWM [41]. Los resultados mostraron, en general, distancias interfascículo bajas y altos porcentajes de intersección para los fascículos simulados. Se realizó una comparativa entre ambos simuladores, donde los mejores resultados los obtuvo el simulador que usa curvas *splines*. Este simulador en su versión inicial, utilizando *splines* y un modelo tubular, tiene la limitación de funcionar mejor con fascículos que presenten formas cilíndricas. No obstante, con las modificaciones introducidas para la evaluación del algoritmo de parcelación cortical, se logró simular una mayor variedad de fascículos con extremos más complejos.

El marco propuesto para la simulación de fascículos de fibras es fácil de usar ya que cuenta con un número razonable de parámetros y el código disponible es intuitivo para el usuario. El simulador de fascículos puede emplearse tanto para la evaluación de algoritmos de *clustering* de fibras como para la parcelación cortical basada en conectividad estructural. Los experimentos realizados han demostrado la aplicabilidad del simulador y de las métricas utilizadas. A pesar de sus limitaciones, el simulador propuesto aborda una importante carencia de datos de referencia para la validación de estos algoritmos. Se considera que esta herramienta podría ser utilizada por la comunidad científica para probar y mejorar sus algoritmos de agrupamiento de fibras y parcelación cortical basada en conectividad estructural. De ahí la relevancia y la necesidad de los algoritmos presentados en esta tesis para la comunidad científica y clínica, dado que los algoritmos validados con el

simulador son de gran importancia en el ámbito de la neurociencia. Además, se pone a disposición de la comunidad científica la base de datos de parcelaciones y sus conexiones simuladas para ser utilizadas en estudios propios.

8.1. Trabajo Futuro

Como trabajo futuro, puede aumentarse la complejidad de la forma de los fascículos al especificar diferentes parámetros como densidad, porcentaje de traslape entre conexiones, distribución de densidad de fibras en cada conexión, conexiones faltantes, entre otros factores.

Dado que el simulador se basa en curvas matemáticas, puede utilizarse para representar fibras de cualquier base de datos. Un trabajo futuro interesante sería probar su funcionamiento con fibras de bases de datos como el *Human Connectome Project* (HCP) [77], demostrando así la versatilidad del simulador.

Otro posible desarrollo futuro del simulador es su uso para simular formas de fibras ruidosas, que son especialmente complejas y variadas, y pueden afectar significativamente los resultados de los algoritmos de análisis de tractografía. Esta capacidad sería valiosa para validar algoritmos de filtrado de fibras, ya que permitiría añadir fibras ruidosas simuladas a un fascículo simulado. De este modo, se podría evaluar la efectividad de los algoritmos de filtrado en la eliminación de ruido.

Finalmente, en el simulador para validar algoritmos de parcelación cortical, resultaría valioso incorporar regiones subcorticales. Además, se podría explorar la generación de conexiones simuladas utilizando otros tipos de parcelaciones corticales, como aquellas basadas en atlas anatómicos, ya que este tipo de parcelaciones ofrece una representación más realista. Otro aspecto importante a desarrollar sería simular variaciones anatómicas entre sujetos para evaluar cómo

los algoritmos de parcelación responden a estos cambios.

8.2. Publicaciones

Durante el desarrollo de esta tesis se publicaron los siguientes trabajos:

8.2.1. Conferencias

Elida Poo, Carlos Kublik, Josselin Houenou, Cyril Poupon, Jean-François Mangin, Cecilia Hernández, Pamela Guevara, "Fiber bundles simulator using exponential curves to validate fiber clustering algorithms," Proc. SPIE 12567, 18th International Symposium on Medical Information Processing and Analysis, 125670J (6 March 2023); <https://doi.org/10.1117/12.2669811>.

En este trabajo se presenta el simulador basado en curvas exponenciales. Se realiza la validación del simulador usando un DWM atlas. Se muestra la aplicabilidad del simulador para la validación de un algoritmo de *clustering* de fibras ampliamente usado en el estado del arte *QuickBundles*. El trabajo fue presentado en 18th International Symposium on Medical Information Processing and Analysis e indexado en scopus.

Elida Poo, Jean-François Mangin, Cecilia Hernández, Pamela Guevara. (2023). BRAIN FIBER BUNDLE SIMULATOR USING SPLINE CURVES. IBRO Neuroscience Reports. 15. S783. 10.1016/j.ibneur.2023.08.1611.

En este trabajo se presenta la primera aproximación al simulador que usa curvas *splines* para la representación de la fibra. Este trabajo fue un abstract y fue publicado en Neuroscience Reports de la conferencia International Brain Research Organization (IBRO).

8.2.2. Revistas WOS

Poo E, Mangin JF, Poupon C, Hernández C, Guevara P. PhyberSIM: a tool for the generation of ground truth to evaluate brain fiber clustering algorithms. *Front Neurosci.* 2024 May 30;18:1396518. doi: 10.3389/fnins.2024.1396518. PMID: 38872943; PMCID: PMC11169570.

En este trabajo se presenta el simulador usando curvas *splines* para la representación de las fibras. El simulador es validado con respecto a un atlas de fibras, además se realiza una comparación con los resultados obtenidos del simulador previo que usa curvas exponenciales. La aplicabilidad del simulador se mostró en la validación de dos algoritmos de *clustering* existentes en el estado del arte *QuickBundles* y *FFclust*. Este trabajo fue publicado en la revista *Frontiers in Neuroscience*.

Poo E, Molina J, Mangin JF, Hernández C, Guevara P. A simulator for the validation of tractography-based cortical surface parcellations.[Por enviar a *Frontiers in Neuroinformatics*].

La investigación presenta una extensión del simulador usando curvas *splines*. Esta extensión modela un fascículo en los extremos según la forma de una parcela cortical mientras que las regiones intermedias y centrales del fascículo tienen una forma tubular. Con el simulador se generan tractografías de cerebro completo para la validación de un algoritmo de parcelación cortical basado en conectividad estructural.

Bibliografía

- [1] Anwander, A., Tittgemeyer, M., Cramon, D., Friederici, A., and Knösche, T. (2007). Connectivity-based parcellation of broca’s area. *Cerebral Cortex*, 17:816–25.
- [2] Arslan, S., Ktena, S. I., Makropoulos, A., Robinson, E. C., Rueckert, D., and Parisot, S. (2018). Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *NeuroImage*, 170:5–30. Segmenting the Brain.
- [3] Basser, P. J., Mattiello, J., and LeBihan, D. (1994). Mr diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66:259–67.
- [4] Batchelor, P. G., Calamante, F., Tournier, J.-D., Atkinson, D., Hill, D. L. G., and Connelly, A. (2006). Quantification of the shape of fiber tracts. *Magnetic Resonance in Medicine*, 55:894–903.
- [5] Behrens, T., Johansen-Berg, H., Woolrich, M., Smith, S., Gandini Wheeler-Kingshott, C., Boulby, P., Barker, G., Sillery, E., Sheehan, K., Ciccarelli, O., Thompson, A., Brady, M., and Matthews, P. (2003). Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, 6:750–7.
- [6] Brady Jr., R. O., Tandon, N., Masters, G. A., Margolis, A., Cohen, B. M., Keshavan, M., and Öngür, D. (2017). Differential brain network activity across mood states in bipolar disorder. *Journal of Affective Disorders*, 207:367–376.
- [7] Brohée, S. and van Helden, J. (2006). Evaluation of clustering algorithms for protein interaction networks. *BMC Bioinformatics*, 7:488.
- [8] Chowdhury, R. and Ramachandran, V. (2018). Cache-oblivious buffer heap and cache-efficient computation of shortest paths in graphs. *ACM Transactions on Algorithms*, 14:1–33.
- [9] Close, T. G., Tournier, J.-D., Calamante, F., Johnston, L. A., Mareels, I., and Connelly, A. (2009). A software tool to generate simulated white matter structures for the assessment of fibre-tracking algorithms. *NeuroImage*, 47(4):1288–1300.
- [10] Cohen, E., Riesenfeld, R., and Elber, G. (2001). *Geometric Modeling with Splines: An Introduction*.

- [11] Colby, J. B., Soderberg, L., Lebel, C., Dinov, I. D., Thompson, P. M., and Sowell, E. R. (2012). Along-tract statistics allow for enhanced tractography analysis. *NeuroImage*, 59(4):3227–3242.
- [12] Craddock, C., James, A., Holtzheimer, P., Hu, X., and Mayberg, H. (2012). A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human brain mapping*, 33:1914–28.
- [13] De Reuck, J. (2014). White matter. In Aminoff, M. J. and Daroff, R. B., editors, *Encyclopedia of the Neurological Sciences (Second Edition)*, pages 759–762. Academic Press, Oxford, second edition edition.
- [14] de Reus, M. A. and van den Heuvel, M. P. (2013). The parcellation-based connectome: Limitations and extensions. *NeuroImage*, 80:397–404. Mapping the Connectome.
- [15] Dennis, E. L., Jahanshad, N., McMahon, K. L., de Zubicaray, G. I., Martin, N. G., Hickie, I. B., Toga, A. W., Wright, M. J., and Thompson, P. M. (2013). Development of brain structural connectivity between ages 12 and 30: A 4-tesla diffusion imaging study in 439 adolescents and adults. *NeuroImage*, 64:671–684.
- [16] Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007a). Regularized, fast, and robust analytical q-ball imaging. *Magnetic Resonance in Medicine*, 58:497–510.
- [17] Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007b). Regularized, fast, and robust analytical q-ball imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(3):497–510.
- [18] Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980.
- [19] Dijkstra, E. (2022). *A Note on Two Problems in Connexion with Graphs*, pages 287–290.
- [20] Duclap, D., Lebois, A., Schmitt, B., Riff, O., Guevara, P., Marrakchi-Kacem, L., Brion, V., Poupon, F., Mangin, J., and Poupon, C. (2012). Connectomist-2.0: a novel diffusion analysis toolbox for brainvisa. In *Proceedings of the 29th ESMRMB meeting*, volume 842.
- [21] Durrleman, S., Fillard, P., Pennec, X., Trouvé, A., and Ayache, N. (2011). Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents. *NeuroImage*, 55(3):1073–1090.
- [22] Feng, Y., Chandio, B. Q., Villalón-Reina, J. E., Thomopoulos, S. I., Joshi, H., Nair, G., Joshi, A. A., Venkatasubramanian, G., John, J. P., and Thompson, P. M. (2023). BundleCleaner: Unsupervised Denoising and Subsampling of

- Diffusion MRI-Derived Tractography Data. In *Computational Diffusion MRI*, pages 152–164.
- [23] Fields, R. (2010). Change in the brain’s white matter. *Science (New York, N.Y.)*, 330:768–9.
- [24] Fillard, P., Descoteaux, M., Goh, A., Gouttard, S., Jeurissen, B., Malcolm, J., Ramirez-Manzanares, A., Reisert, M., Sakaie, K., Tensaouti, F., Yo, T., Mangin, J.-F., and Poupon, C. (2011). Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom. *NeuroImage*, 56:220–34.
- [25] Fisher, R. and Yates, F. (1963). *Statistical Tables for Biological, Agricultural and Medical Research*. Longman, 6th edition.
- [26] Frigo, M., Costantini, I., Deriche, R., and Deslauriers-Gauthier, S. (2019). *Resolving the Crossing/Kissing Fiber Ambiguity Using Functionally Informed COMMIT*, pages 335–343.
- [27] Gallardo, G., Wells, W., Deriche, R., and Wassermann, D. (2018). Groupwise structural parcellation of the whole cortex: A logistic random effects model based approach. *NeuroImage*, 170:307–320. Segmenting the Brain.
- [28] Gallardo Diez, G. A. (2018). *Inferring and comparing structural parcellations of the human brain using diffusion MRI*. Theses, COMUE Université Côte d’Azur (2015 - 2019).
- [29] Garyfallidis, E., Brett, M., Correia, M. M., Williams, G. B., and Nimmo-Smith, I. (2012). Quickbundles, a method for tractography simplification. *Frontiers in neuroscience*, 6:175.
- [30] Garyfallidis, E., Côté, M.-A., Rheault, F., and Descoteaux, M. (2016). QuickBundlesX: Sequential clustering of millions of streamlines in multiple levels of detail at record execution time. In *International Society of Magnetic Resonance in Medicine Conference (ISMRM)*.
- [31] Garyfallidis, E., Côté, M.-A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., and Descoteaux, M. (2018). Recognition of white matter bundles using local and global streamline-based registration and clustering. *NeuroImage*, 170:283–295. Segmenting the Brain.
- [32] Glozman, T., Bruckert, L., Pestilli, F., Yecies, D. W., Guibas, L. J., and Yeom, K. W. (2018). Framework for shape analysis of white matter fiber bundles. *NeuroImage*, 167:466–477.
- [33] Goicovich, I., Olivares, P., Román, C., Vázquez, A., Poupon, C., Mangin, J.-F., Guevara, P., and Hernández, C. (2021). Fiber clustering acceleration with a modified kmeans++ algorithm using data parallelism. *Frontiers in Neuroinformatics*, 15.
- [34] Gopinath, K., Desrosiers, C., and Lombaert, H. (2019). Graph convolutions

- on spectral embeddings for cortical surface parcellation. *Medical Image Analysis*, 54:297–305.
- [35] Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., and Petersen, S. E. (2014). Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cerebral Cortex*, 26(1):288–303.
- [36] Grover, V. P., Tognarelli, J. M., Crossey, M. M., Cox, I. J., Taylor-Robinson, S. D., and McPhail, M. J. (2015). Magnetic resonance imaging: Principles and techniques: Lessons for clinicians. *Journal of Clinical and Experimental Hepatology*, 5(3):246–255.
- [37] Guevara, M., Román, C., Houenou, J., Duclap, D., Poupon, C., Mangin, J. F., and Guevara, P. (2017). Reproducibility of superficial white matter tracts using diffusion-weighted imaging tractography. *Neuroimage*, 147:703–725.
- [38] Guevara, P., Duclap, D., Marrakchi-Kacem, L., Rivière, D., Cointepas, Y., Poupon, C., and Mangin, J.-F. (2011a). Accurate tractography propagation mask using t1-weighted data rather than fa. *Proc. Intl. Soc. Mag. Reson. Med. (ISMRM)*, 19.
- [39] Guevara, P., Duclap, D., Poupon, C., Marrakchi-Kacem, L., Fillard, P., Le Bihan, D., Leboyer, M., Houenou, J., and Mangin, J.-F. (2012). Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas. *NeuroImage*, 61:1083–99.
- [40] Guevara, P., Poupon, C., Rivière, D., Cointepas, Y., Descoteaux, M., Thirion, B., and Mangin, J.-F. (2011b). Robust clustering of massive tractography datasets. *NeuroImage*, 54(3):1975–1993.
- [41] Guevara Alvez, P. B. (2011). *Inference of a human brain fiber bundle atlas from high angular resolution diffusion imaging*. Theses, Université Paris Sud - Paris XI.
- [42] Hernandez, C., Mella, C., Navarro, G., Olivera-Nappa, A., and Araya, J. (2017). Protein complex prediction via dense subgraphs and false positive analysis. *PLOS ONE*, 12:1–37.
- [43] Hinrichs, A., Krieg, D., Kunsch, R. J., and Rudolf, D. (2020). Expected dispersion of uniformly distributed points.
- [44] Huerta, I., Vázquez, A., López-López, N., Houenou, J., Poupon, C., Mangin, J.-F., Guevara, P., and Hernández, C. (2020). Inter-subject clustering of brain fibers from whole-brain tractography. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1687–1691.
- [45] Jeurissen, B., Descoteaux, M., Mori, S., and Leemans, A. (2019). Diffusion mri fiber tractography of the brain. *NMR in Biomedicine*, 32(4):e3785.
- [46] Ji, J., Zhang, A., Liu, C., Quan, X., and Liu, Z. (2014). Survey: Functional

- module detection from protein-protein interaction networks. *Knowledge and Data Engineering, IEEE Transactions on*, 26:261–277.
- [47] Labra Avila, N., Guevara, P., Duclap, D., Houenou, J., Poupon, C., Mangin, J.-F., and Figueroa, M. (2017). Fast automatic segmentation of white matter streamlines based on a multi-subject bundle atlas. *Neuroinformatics*, 15.
- [48] Le Bihan, D. and Iima, M. (2015). Diffusion magnetic resonance imaging: What water tells us about biological tissues. *PLOS Biology*, 13:1–13.
- [49] Lefranc, S., Roca, P., Perrot, M., Poupon, C., Le Bihan, D., Mangin, J.-F., and Rivière, D. (2016). Groupwise connectivity-based parcellation of the whole human cortical surface using watershed-driven dimension reduction. *Medical Image Analysis*, 30:11–29.
- [50] Legarreta, J. H., Petit, L., Jodoin, P.-M., and Descoteaux, M. (2022). Clustering in Tractography Using Autoencoders (CINTA). In *Computational Diffusion MRI: 13th International Workshop, CDMRI 2022, Held in Conjunction with MICCAI 2022, Singapore, Singapore, September 22, 2022, Proceedings*.
- [51] Losnegård, A., Lundervold, A., and Hodneland, E. (2013). White matter fiber tracking directed by interpolating splines and a methodological framework for evaluation. *Frontiers in Neuroinformatics*, 7.
- [52] López-López, N., Vázquez, A., Houenou, J., Poupon, C., Mangin, J.-F., Ladra, S., and Guevara, P. (2020a). From coarse to fine-grained parcellation of the cortical surface using a fiber-bundle atlas. *Frontiers in Neuroinformatics*, 14:32.
- [53] López-López, N., Vázquez, A., Poupon, C., Mangin, J.-F., Ladra, S., and Guevara, P. (2020b). Geosp: A parallel method for a cortical surface parcellation based on geodesic distance. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1696–1700.
- [54] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297.
- [55] Maier-Hein, K., Neher, P., Houde, J.-C., Côté, M.-A., Garyfallidis, E., Zhong, J., Chamberland, M., Yeh, F.-C., Lin, Y.-C., Ji, Q., Reddick, W., Glass, J., Chen, D., Yuanjing, F., Gao, C., Wu, Y., Ma, J., Renjie, H., Li, Q., and Descoteaux, M. (2017). The challenge of mapping the human connectome based on diffusion tractography. *Nature Communications*, 8.
- [56] McRobbie, D. W., Moore, E. A., Graves, M. J., and Prince, M. R. (2006). *MRI from Picture to Proton*. Cambridge University Press, 2 edition.
- [57] Mendoza, C., Román, C., Mangin, J.-F., Hernández, C., and Guevara, P. (2024). Short fiber bundle filtering and test-retest reproducibility of the superficial white matter. *Frontiers in Neuroscience*, 18.
- [58] Mimenza, A. J., Navarro, S. G. A., Albán, P. B., Samudio, A., Ortega,

- J. R., Meléndez, O. M., and Funes, J. A. A. (2020). Diffusion tensor imaging (tractography) in elderly people with mixed dementia and mild Alzheimer’s disease. *Alzheimer’s & Dementia*, 16(S5):e041903.
- [59] Moberts, B., Vilanova, A., and van Wijk, J. (2005). Evaluation of fiber clustering methods for diffusion tensor imaging. In *VIS 05. IEEE Visualization, 2005.*, pages 65–72.
- [60] Moghimi, P., Dang, A. T., Netoff, T. I., Lim, K. O., and Atluri, G. (2021). A review on mr based human brain parcellation methods.
- [61] Molina, J., Mendoza, C., Román, C., Houenou, J., Poupon, C., Mangin, J. F., El-Deredy, W., Hernández, C., and Guevara, P. (2023). Group-wise cortical parcellation based on structural connectivity and hierarchical clustering. In *18th International Symposium on Medical Information Processing and Analysis*, volume 12567, page 125670L.
- [62] Möller, T. and Trumbore, B. (2005). Fast, minimum storage ray/triangle intersection. In *ACM SIGGRAPH 2005 Courses*, SIGGRAPH ’05, page 7–es, New York, NY, USA. Association for Computing Machinery.
- [63] Moreno-Dominguez, D. (2014). Whole-brain cortical parcellation: A hierarchical method based on dmri tractography. *Max Planck Institute for Human Cognitive and BrainSciences, 2014*, pages 816–25.
- [64] Mori, S. and van Zijl, P. C. M. (2002). Fiber tracking: principles and strategies - a technical review. *NMR in biomedicine*, 15:468–80.
- [65] Mueller, J. L., Stavelly, R., Hotta, R., and Goldstein, A. M. (2022). Peripheral nervous system: A promising source of neuronal progenitors for central nervous system repair. *Frontiers in Neuroscience*, 16.
- [66] Mukherjee, P., Berman, J. I., Chung, S., Hess, C. P., and Henry, R. G. (2008). Diffusion tensor mr imaging and fiber tractography: Theoretic underpinnings. *American Journal of Neuroradiology*, 29:632 – 641.
- [67] Neher, P., Laun, F., Stieltjes, B., and Maier-Hein, K. (2014). Fiberfox: Facilitating the creation of realistic white matter software phantoms. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 72.
- [68] Nepusz, T., Yu, H., and Paccanaro, A. (2012). Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods*, 9:471–2.
- [69] O’Donnell, L., Kubicki, M., Shenton, M., Dreusicke, M., Grimson, W., and Westin, C.-F. (2006). A Method for Clustering White Matter Fiber Tracts. *AJNR. American Journal of Neuroradiology*, 27:1032–6.
- [70] O’Donnell, L. J. and Westin, C.-F. (2007). Automatic tractography segmentation using a high-dimensional white matter atlas. *IEEE Transactions on Medical Imaging*, 26(11):1562–1575.

- [71] Pardo, E., Guevara, P., Duclap, D., Houenou, J., Lebois, A., Schmitt, B., Le Bihan, D., Mangin, J.-F., and Poupon, C. (2013). Study of the variability of short association bundles on a hardi database. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 77–80.
- [72] Poo, E., Kublik, C., Houenou, J., Poupon, C., Mangin, J.-F., Hernández, C., and Guevara, P. (2023). Fiber bundles simulator using exponential curves to validate fiber clustering algorithms. In *18th International Symposium on Medical Information Processing and Analysis*, volume 12567, page 125670J.
- [73] Poo, E., Mangin, J.-F., Poupon, C., Hernández, C., and Guevara, P. (2024). Phybersim: a tool for the generation of ground truth to evaluate brain fiber clustering algorithms. *Frontiers in Neuroscience*, 18.
- [74] Poupon, C., Laribiere, L., Tournier, G., Bernard, J., Fournier, D., Fillard, P., Descoteaux, M., and Mangin, J.-F. (2010). A diffusion hardware phantom looking like a coronal brain slice. In *ISMRM 18th Scientific Meeting and Exhibition*, Stockholm, Sweden.
- [75] Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., and Mangin, J.-F. (2008). New diffusion phantoms dedicated to the study and validation of high-angular-resolution diffusion imaging (hardi) models. *Magnetic Resonance in Medicine*, 60:1276–83.
- [76] Prieto, Y., Molina, J., Otero, M., Mangin, J.-F., Hernández, C., El-Dereby, W., and Guevara, P. (2023). Multiscale cortical parcellation based on geodesic distance and hierarchical clustering. In *2023 19th International Symposium on Medical Information Processing and Analysis (SIPAIM)*, pages 1–5.
- [77] Project, H. C. (n.d.). Hcp young adult, 1200 subjects data release. Online. <https://www.humanconnectome.org/study/hcp-young-adult/document/1200-subjects-data-release>.
- [78] Reichenbach, A., Goldau, M., Heine, C., and Hlawitschka, M. (2015). V-bundles: Clustering fiber trajectories from diffusion mri in linear time. *Medical Image Computing and Computer Assisted Intervention*, 9349:191–198.
- [79] Rodrigues (1840). Des lois géométriques qui régissent les déplacements d’un système solide dans l’espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *Journal de Mathématiques Pures et Appliquées*, 98:380–440.
- [80] Román, C., Guevara, M., Valenzuela, R., Figueroa, M., Houenou, J., Duclap, D., Poupon, C., Mangin, J.-F., and Guevara, P. (2017). Clustering of whole-brain white matter short association bundles using hardi data. *Frontiers in Neuroinformatics*, 11:73.
- [81] Román, C., Hernández, C., Figueroa, M., Houenou, J., Poupon, C., Mangin, J.-F., and Guevara, P. (2022). Superficial white matter bundle atlas based on

- hierarchical fiber clustering over probabilistic tractography data. *NeuroImage*, 262:119550.
- [82] Salehi, M., Karbasi, A., Shen, X., Scheinost, D., and Constable, R. (2017). An exemplar-based approach to individualized parcellation reveals the need for sex specific functional networks. *NeuroImage*, 170.
- [83] Schmahmann, J. D. and Pandya, D. N. (2006). *Fiber Pathways of the Brain*. Oxford University Press.
- [84] Schmitt, B., Lebois, A., Duclap, D., Guevara, P., Poupon, F., Rivière, D., Cointepas, Y., LeBihan, D., Mangin, J., and Poupon, C. (2012). Connect/archi: an open database to infer atlases of the human brain connectivity. *ESMRMB*, 272:2012.
- [85] Sculley, D. (2010). Web-scale k-means clustering. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, page 1177–1178, New York, NY, USA. Association for Computing Machinery.
- [86] Sheng, J., Shao, M., Zhang, Q., Zhou, R., Wang, L., and Xin, Y. (2020). Alzheimer’s disease, mild cognitive impairment, and normal aging distinguished by multi-modal parcellation and machine learning. *Scientific Reports*, 10.
- [87] Silva, F., Guevara, M., Poupon, C., Mangin, J.-F., Hernandez, C., and Guevara, P. (2019). Cortical surface parcellation based on graph representation of short fiber bundle connections. pages 1479–1482.
- [88] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, 1.
- [89] Stiles, J. and Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology Review*, 20:327 – 348.
- [90] Taylor, C. J. and Kriegman, D. J. (1994). Minimization on the lie group $so(3)$ and related manifolds. In *Yale Center For Systems Science Technical Report*.
- [91] Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion mri: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage*, 35(4):1459–1472.
- [92] Urchs, S., Armoza, J., Moreau, C., Benhajali, Y., St-Aubin, J., Orban, P., and Bellec, P. (2019). Mist: A multi-resolution parcellation of functional brain networks. *MNI Open Research*, 1:3.
- [93] Van Loock, W., Pipeleers, G., and Swevers, J. (2015). B-spline parameterized optimal motion trajectories for robotic systems with guaranteed constraint satisfaction. *Mechanical Sciences*, 6:163–171.
- [94] Vergara, C., Silva, F., Huerta, I., López-López, N., Vázquez, A., Houenou, J., Poupon, C., Mangin, J.-F., Hernández, C., and Guevara, P. (2021). Group-wise cortical surface parcellation based on inter-subject fiber clustering. In *2021*

- 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 2655–2659.
- [95] Vázquez, A., López-López, N., Labra, N., Figueroa, M., Poupon, C., Mangin, J.-F., Hernández, C., and Guevara, P. (2019). Parallel optimization of fiber bundle segmentation for massive tractography datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 178–181.
- [96] Vázquez, A., López-López, N., Sánchez, A., Houenou, J., Poupon, C., Mangin, J.-F., Hernández, C., and Guevara, P. (2020). Ffclust: Fast fiber clustering for large tractography datasets for a detailed study of brain connectivity. *NeuroImage*, 220:117070.
- [97] Wang, L., Lin, F. V., Cole, M., and Zhang, Z. (2021). Learning clique subgraphs in structural brain network classification with application to crystallized cognition. *NeuroImage*, 225:117493.
- [98] Webb, W. G. (2017). 2 - organization of the nervous system i. In Webb, W. G., editor, *Neurology for the Speech-Language Pathologist (Sixth Edition)*, pages 13–43. Mosby, sixth edition edition.
- [99] Wu, X., Bi, W., Zhu, J., Yang, L., and Xie, M. (2009). Optimized white matter fiber reconstruction with b-spline curve and evolutionary computation. In *2009 Fifth International Conference on Natural Computation*, volume 4, pages 384–387.
- [100] Xu, D. and Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2.
- [101] Yang, J. Y.-M., Yeh, C.-H., Poupon, C., and Calamante, F. (2021). Diffusion MRI tractography for neurosurgery: the basics, current state, technical reliability and challenges. *Physics in Medicine & Biology*, 66(15):15TR01.
- [102] Yeh, F.-C. (2020). Shape analysis of the human association pathways. *NeuroImage*, 223:117329.
- [103] Zhang, F., Daducci, A., He, Y., Schiavi, S., Seguin, C., Smith, R. E., Yeh, C.-H., Zhao, T., and O’Donnell, L. J. (2022). Quantitative mapping of the brain’s structural connectivity using diffusion MRI tractography: A review. *NeuroImage*, 249:118870.

Apéndice A

Material Suplementario

A.1. Simulador de fascículos usando curvas splines

A.1.1. Modelo preliminar utilizando regiones de sección transversal elípticas

Para probar la viabilidad de utilizar un modelo tubular para representar los fascículos, se realizó un breve análisis de la forma de las regiones de los fascículos del atlas, modelándolas con secciones transversales elípticas. Para definir estas regiones elípticas, se requieren tres parámetros esenciales: el radio del eje mayor, el radio del eje menor y el punto central de la elipse. Los puntos centrales y los planos de las elipses son los mismos que se utilizaron para los sectores circulares. Para calcular los radios de cada región, se identifica el punto más alejado del centroide (Pf1) y un segundo punto (Pf2), también distante pero en una dirección perpendicular al primero. Luego, para determinar los radios mayor y menor promedio de cada sección, se calcula la distancia media entre todos los puntos de la región y Pf1 y Pf2, respectivamente, y se divide por dos (ver un ejemplo en la Figura A.1.1). La Figura A.1.2 muestra el modelo obtenido para el fascículo

uncinado izquierdo, comparando las formas elípticas con el modelo tubular circular.

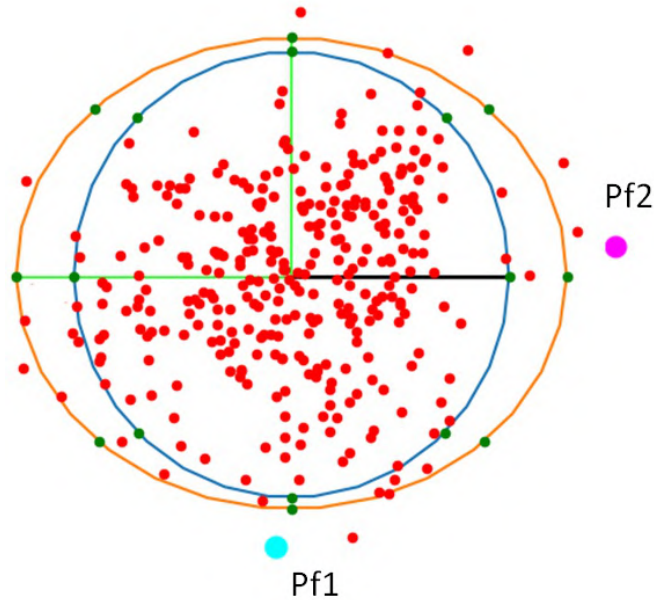


Figura A.1.1: Ejemplo de una representación de una región intermedia de sección transversal del fascículo uncinado izquierdo utilizando el modelo tubular y el modelo preliminar basado en secciones transversales elípticas. Los puntos del fascículo correspondientes a la región están en rojo. La región de sección transversal circular está en azul, mientras que la región de sección transversal elíptica está en naranja. Los puntos periféricos para ambas regiones están en verde. El radio de la representación circular está en negro, y los radios mayor y menor de la elipse están en verde claro. El punto más alejado (Pf1) se muestra en cian, y el punto más alejado en una dirección perpendicular a Pf1 (Pf2) está en magenta.

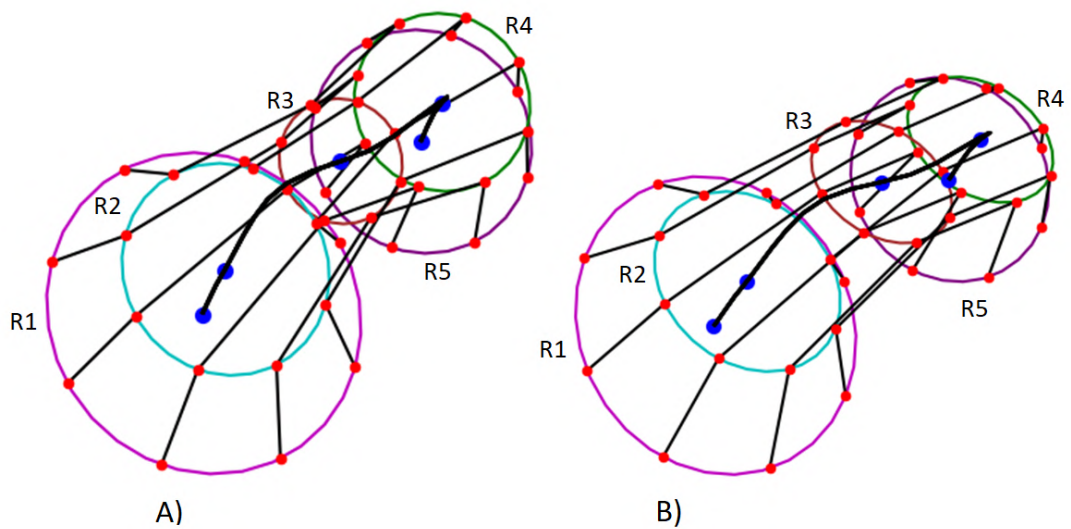


Figura A.1.2: Ejemplo del modelo tubular propuesto (A) junto con un modelo preliminar utilizando secciones transversales elípticas (B) para el fascículo uncinado izquierdo. El centroide del fascículo se muestra con una línea gruesa negra. Las regiones de sección transversal de los extremos aparecen en magenta (RR=0.9) y púrpura (RR=0.9), la región de sección transversal central en marrón (RR=0.7), y las regiones de sección transversal intermedias están en cian (RR=0.8) y verde (RR=0.8). Los puntos centrales de las regiones de sección transversal están en azul, mientras que los puntos periféricos que definen los 8 sectores por región están en rojo. RR: Relación de radios elípticos.

	Radius Ratio					Mean value	Proposed simulator		Preliminary model with elliptical cross-sections	
	R1	R2	R3	R4	R5		ID \pm STD (mm)	IP(%)	ID \pm STD (mm)	IP(%)
Atlas Bundles										
Anterior Left Arcuate	0.8	1.0	0.7	0.9	1.0	0.9	8.2 \pm 1.1	96.4	9.8 \pm 1.5	76.7
Anterior Right Arcuate	0.8	1.0	0.7	0.9	1.0	0.9	8.2 \pm 1.0	97.9	9.9 \pm 1.4	71.2
Left Arcuate	0.8	0.8	1.0	0.6	0.7	0.8	10.5 \pm 2.2	44.8	11.4 \pm 2.8	33.9
Right Arcuate	0.8	0.8	1.0	0.6	0.7	0.8	10.8 \pm 2.2	39.9	11.4 \pm 2.6	34.8
Posterior Left Arcuate	0.9	0.9	0.9	0.8	0.8	0.9	6.6 \pm 1.0	100	7.0 \pm 0.9	100.0
Posterior Right Arcuate	0.9	1.0	0.9	0.8	0.8	0.9	6.7 \pm 0.9	100.0	7.1 \pm 1.0	100.0
Left Cingulum Fibers	1.0	0.8	0.7	0.6	1.0	0.8	7.9 \pm 1.6	86.4	8.6 \pm 1.7	78.2
Right Cingulum Fibers	1.0	0.8	0.7	0.6	1.0	0.8	7.2 \pm 1.4	89.0	8.6 \pm 1.9	78.2
Left Corticospinal Tract	1.0	0.9	0.8	0.9	0.9	0.9	8.1 \pm 1.6	87.3	5.8 \pm 1.3	100.0
Right Corticospinal Tract	1.0	0.9	0.8	0.9	0.9	0.9	8.0 \pm 1.6	87.5	5.9 \pm 1.3	99.8
Left Fornix	0.9	0.8	0.7	0.7	0.5	0.7	10.6 \pm 1.8	32.9	9.5 \pm 1.8	52.9
Right Fornix	0.9	0.8	0.7	0.7	0.5	0.7	10.8 \pm 1.7	30.0	9.6 \pm 1.9	53.6
Left Inferior Fronto-occipital	0.8	0.9	0.7	0.9	0.3	0.7	11.9 \pm 1.7	13.2	12.8 \pm 1.8	11.6
Right Inferior Fronto-occipital	0.8	0.9	0.7	0.9	0.3	0.7	11.5 \pm 1.5	15.5	12.8 \pm 1.7	11.2
Left Inferior Longitudinal	1.0	0.7	0.9	1.0	0.9	0.9	10.3 \pm 1.8	48.2	11.0 \pm 1.8	30.1
Right Inferior Longitudinal	1.0	0.7	0.9	1.0	0.9	0.9	10.2 \pm 1.8	49.6	10.9 \pm 1.7	30.6
Left Frontal Thalamic Radiations	0.7	0.8	0.9	1.0	0.9	0.8	7.9 \pm 1.7	88.5	8.8 \pm 1.9	74.5
Right Frontal Thalamic Radiations	0.7	0.8	0.9	1.0	0.9	0.9	8.1 \pm 1.8	85.4	9.1 \pm 2.1	70.3
Left Superior Motor Thalamic Radiations	0.7	1.0	0.7	0.9	0.8	0.8	7.1 \pm 1.1	99.3	7.3 \pm 1.6	94.3
Right Superior Motor Thalamic Radiations	0.7	1.0	0.7	0.9	0.8	0.8	7.2 \pm 1.3	98.2	7.4 \pm 1.6	95.5
Left Occipital Thalamic Radiations	0.7	0.9	0.8	0.8	0.8	0.8	7.6 \pm 1.6	90.4	7.6 \pm 1.5	96.4
Right Occipital Thalamic Radiations	0.7	0.9	0.8	0.8	0.8	0.8	7.7 \pm 1.5	92.8	7.8 \pm 1.5	96.8
Left Superior Parietal Thalamic Radiations	0.9	0.9	0.6	0.9	0.9	0.9	6.3 \pm 1.1	100.0	6.4 \pm 1.0	100.0
Right Superior Parietal Thalamic Radiations	0.9	0.9	0.6	0.9	0.9	0.8	7.5 \pm 1.2	100.0	6.3 \pm 1.0	100.0
Left Temporal Thalamic Radiations	0.6	1.0	0.7	0.6	0.9	0.8	7.2 \pm 1.6	93.8	9.2 \pm 2.1	61.5
Right Temporal Thalamic Radiations	0.6	0.9	0.7	0.6	0.9	0.8	7.2 \pm 1.4	95.4	9.4 \pm 2.0	63.1
Left Uncinate	0.9	0.8	0.7	0.8	0.9	0.8	8.4 \pm 1.2	90.5	8.1 \pm 1.4	91.4
Right Uncinate	0.9	0.8	0.7	0.8	0.9	0.8	8.6 \pm 1.2	88.6	7.9 \pm 1.4	91.7

Tabla A.1.1: Relación de radios para las cinco regiones de sección transversal elíptica (R1-R5) y relación de radios promedio para los 28 fascículos del atlas de DWM analizados. También se muestran los resultados para el porcentaje de intersección (%) (IP) y la distancia entre fascículos \pm desviación estándar (mm) (ID \pm STD) para el modelo tubular en comparación con el modelo preliminar utilizando secciones transversales elípticas. En azul, se destacan los ocho fascículos que presentan mejoras.

A.1.2. Simulación de fascículos de un atlas de materia blanca profunda

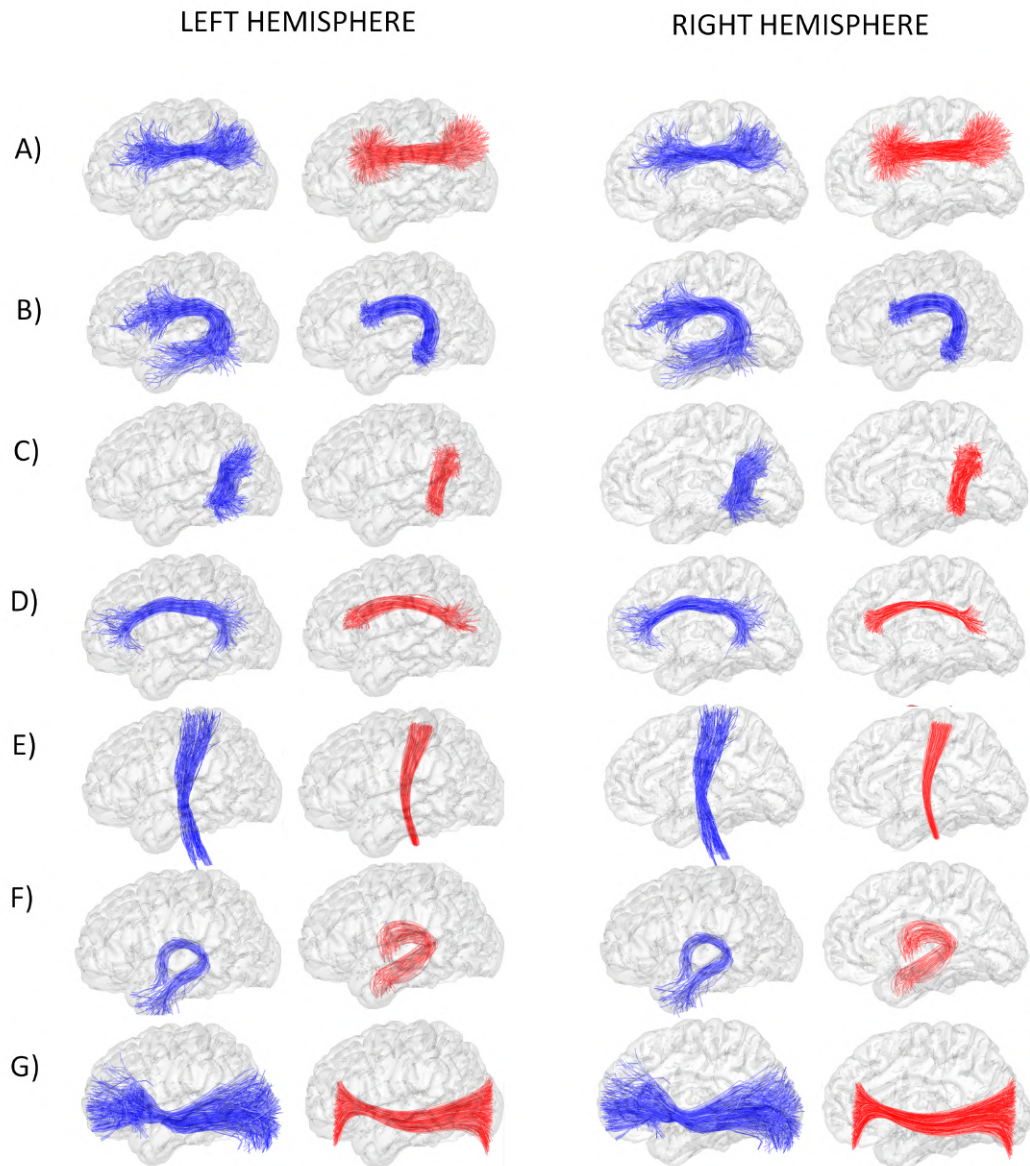


Figura A.1.3: Fascículos simulados del atlas de fascículos DWM (Guevara et al., 2012) para el hemisferio izquierdo (primera columna) y el hemisferio derecho (segunda columna). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo. A) Arqueado anterior, B) Arqueado, C) Arqueado posterior, D) Fibras del cingulum, E) Tracto corticoespinal, F) Fornix, G) Fascículo fronto-occipital inferior.

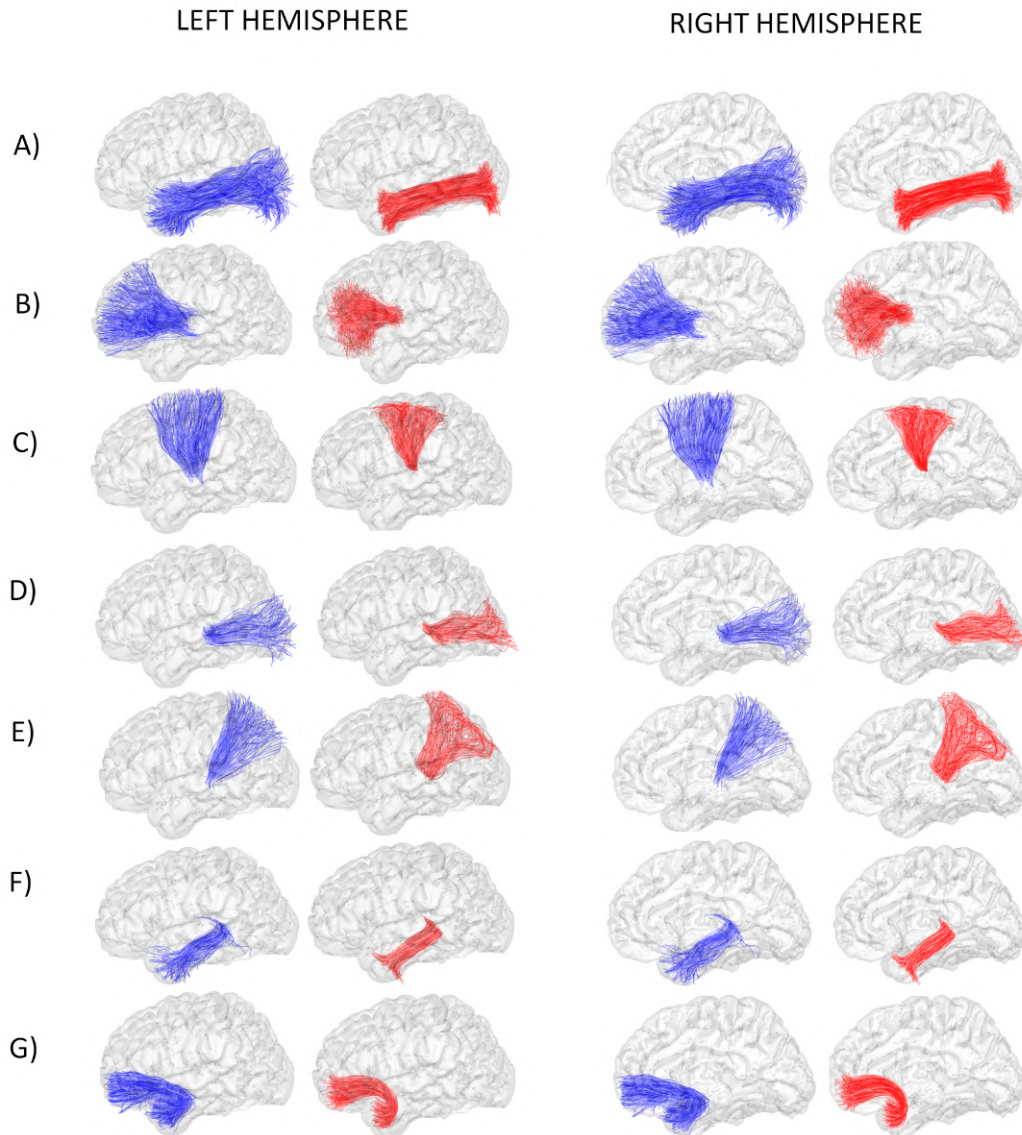


Figura A.1.4: Fascículos simulados del atlas de fascículos DWM (Guevara et al., 2012) para el hemisferio izquierdo (primera columna) y el hemisferio derecho (segunda columna). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo. A) Fascículo longitudinal inferior, B) Radiaciones talámicas frontales, C) Radiaciones talámicas motoras superiores, D) Radiaciones talámicas occipitales, E) Radiaciones talámicas parietales superiores, F) Radiaciones talámicas temporales, G) Fascículo uncinado.

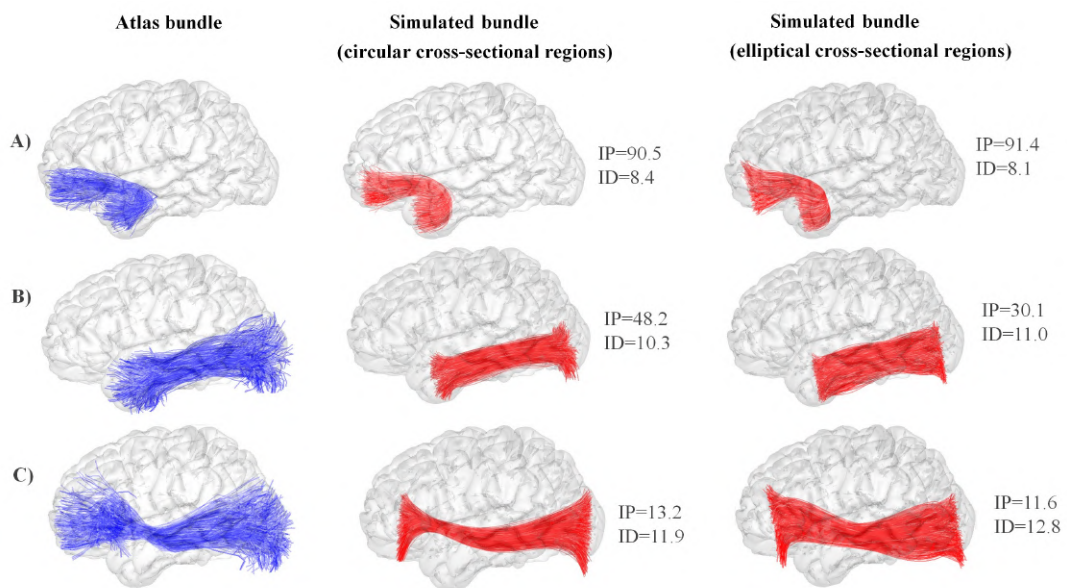


Figura A.1.5: Ejemplos de simulación de fascículos basados en el atlas de fascículos DWM (Guevara et al., 2012). Los fascículos originales del atlas se muestran a la izquierda (en azul), y sus respectivos fascículos simulados se presentan a la derecha en rojo, tanto para el modelo tubular como para el modelo preliminar utilizando secciones transversales elípticas. A) El fascículo uncinado izquierdo, B) El longitudinal inferior izquierdo, y C) El fronto-occipital inferior. IP: Porcentaje de intersección entre fascículos (%) ID: Distancia entre fascículos (mm).

A.1.3. Simulación de conjuntos de datos de cerebro completo

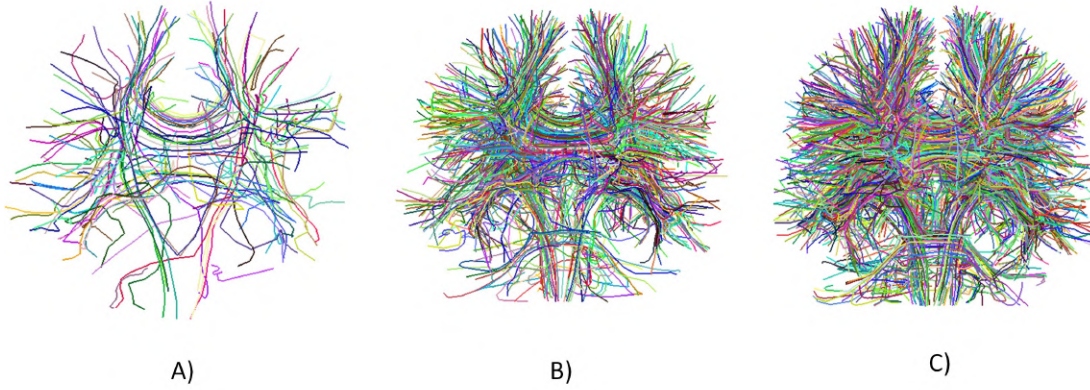


Figura A.1.6: Centroides utilizados para la simulación de los tres conjuntos de datos de referencia de 100 (A), 500 (B) y 1000 (C) fascículos.

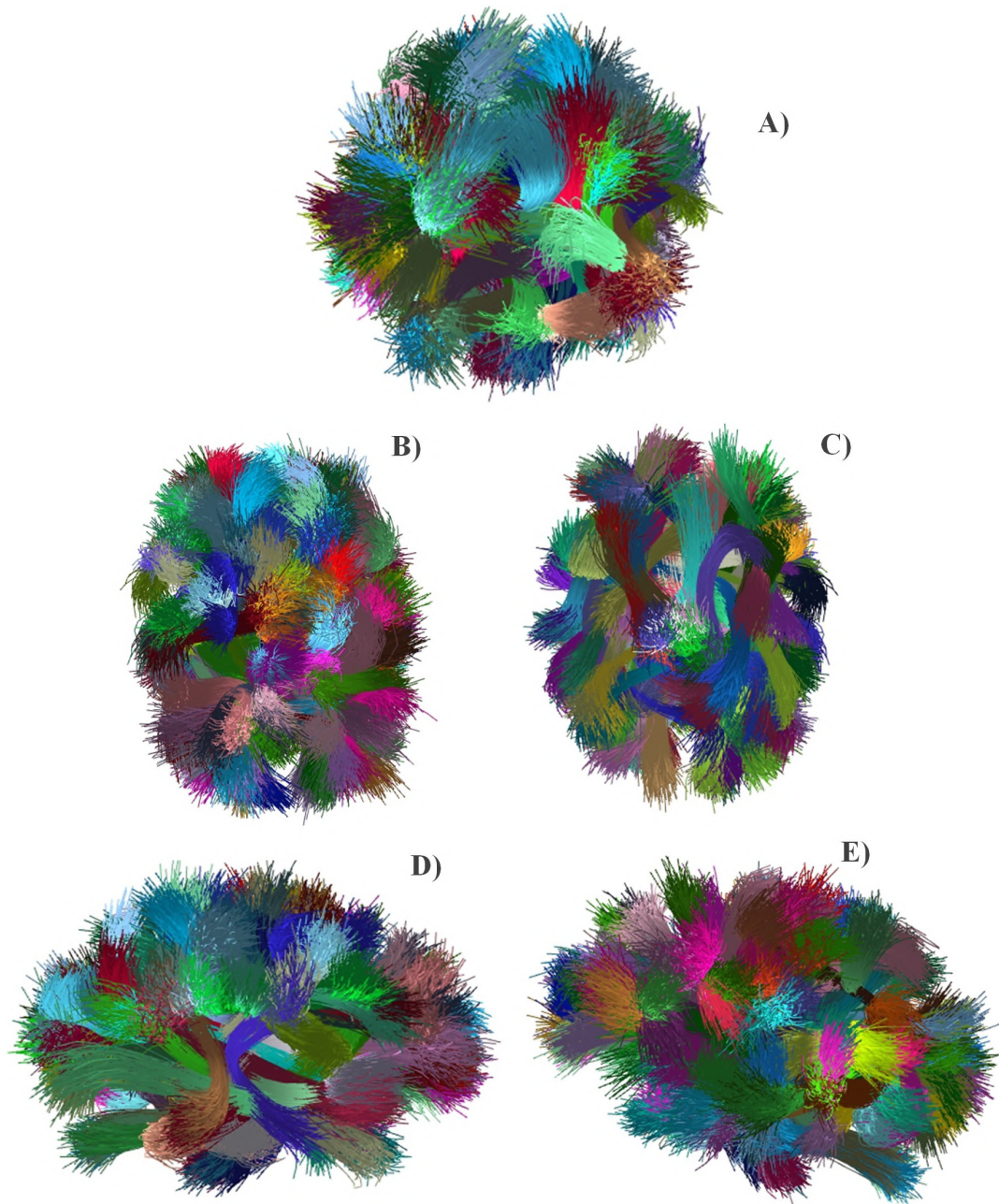


Figura A.1.7: Diferentes vistas del conjunto de datos de referencia de 100 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda, E) Vista lateral derecha.

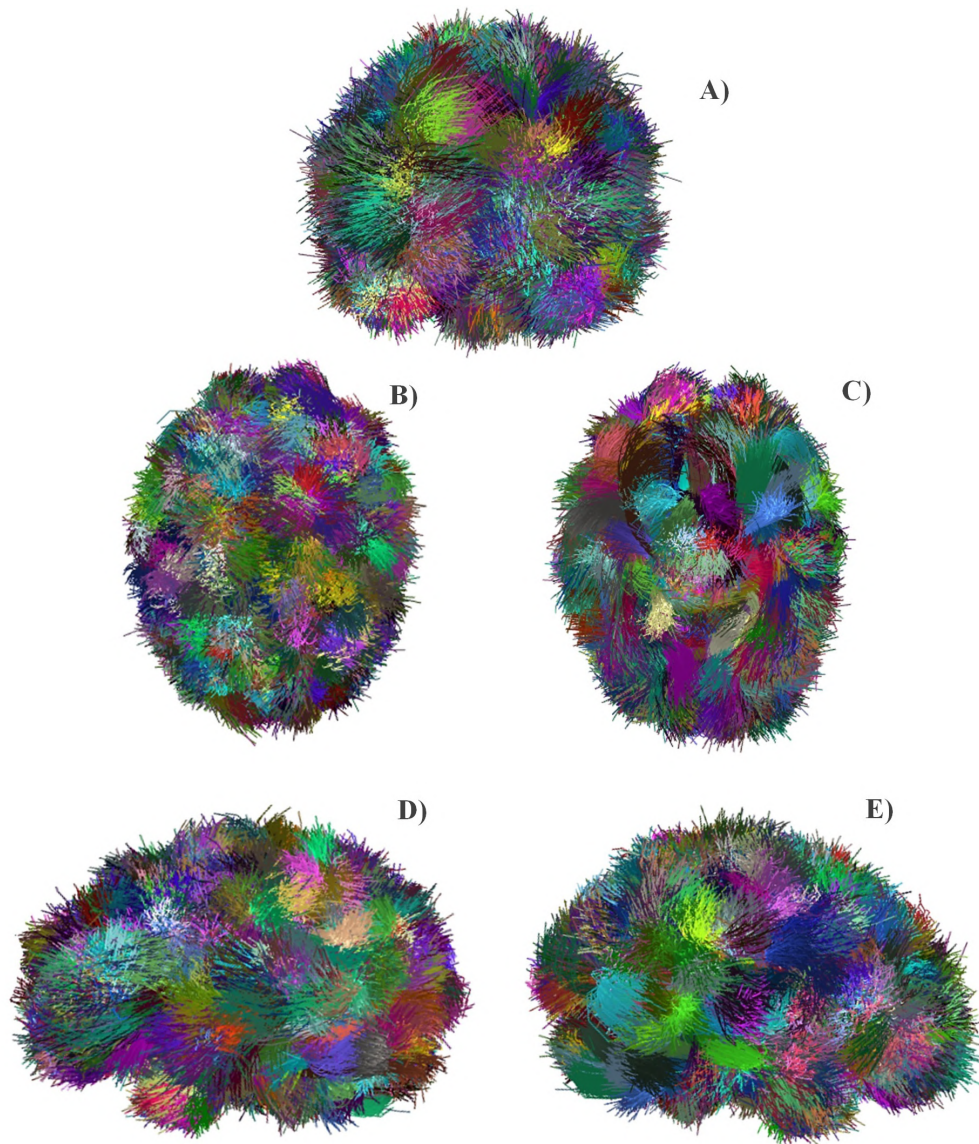


Figura A.1.8: Diferentes vistas del conjunto de datos de referencia de 500 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda y E) Vista lateral derecha.

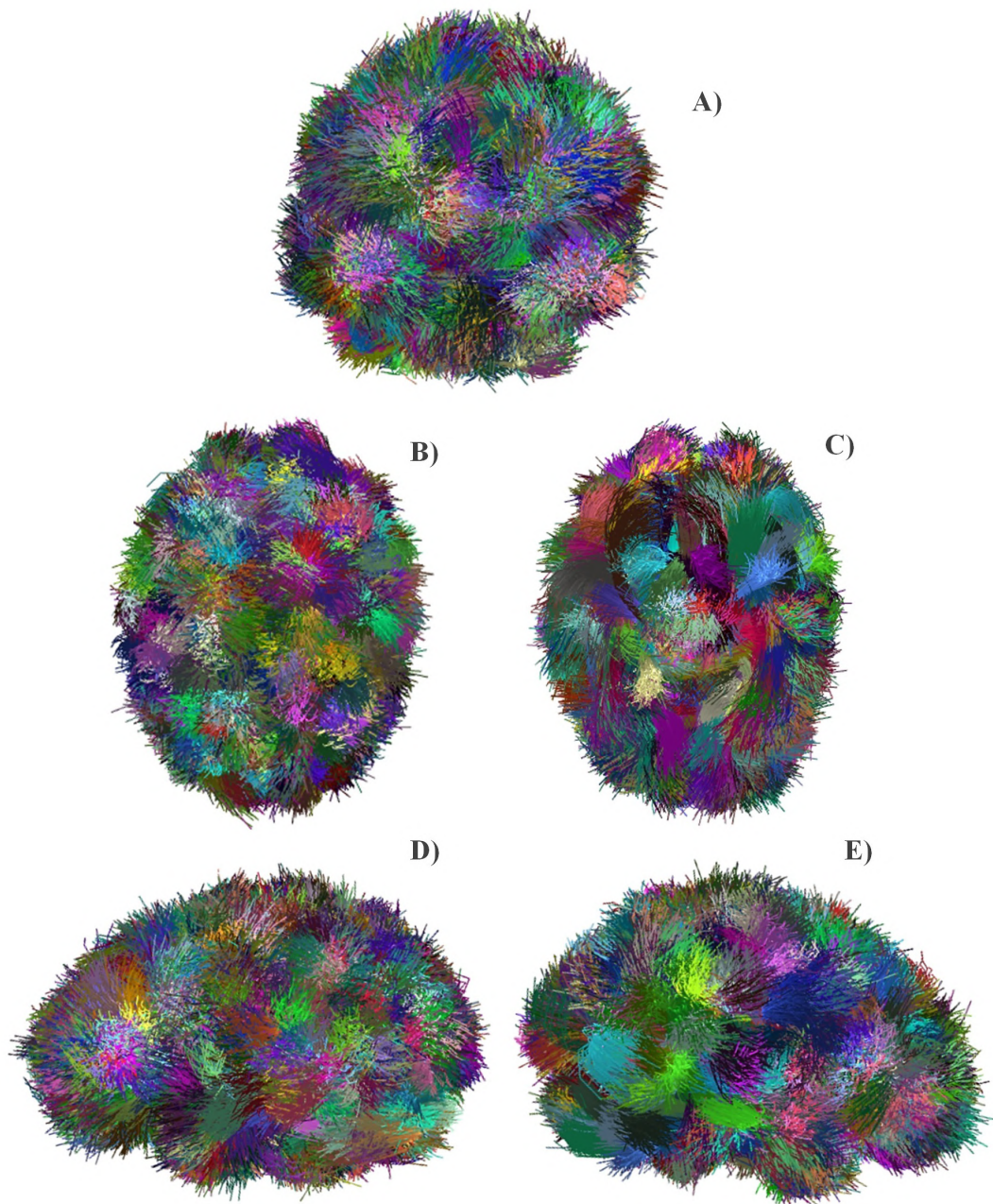


Figura A.1.9: Diferentes vistas del conjunto de datos de referencia de 1000 fascículos: A) Vista coronal, B) Vista superior, C) Vista inferior, D) Vista lateral izquierda y E) Vista lateral derecha.

A.1.4. Cálculo de parámetros de FFClust

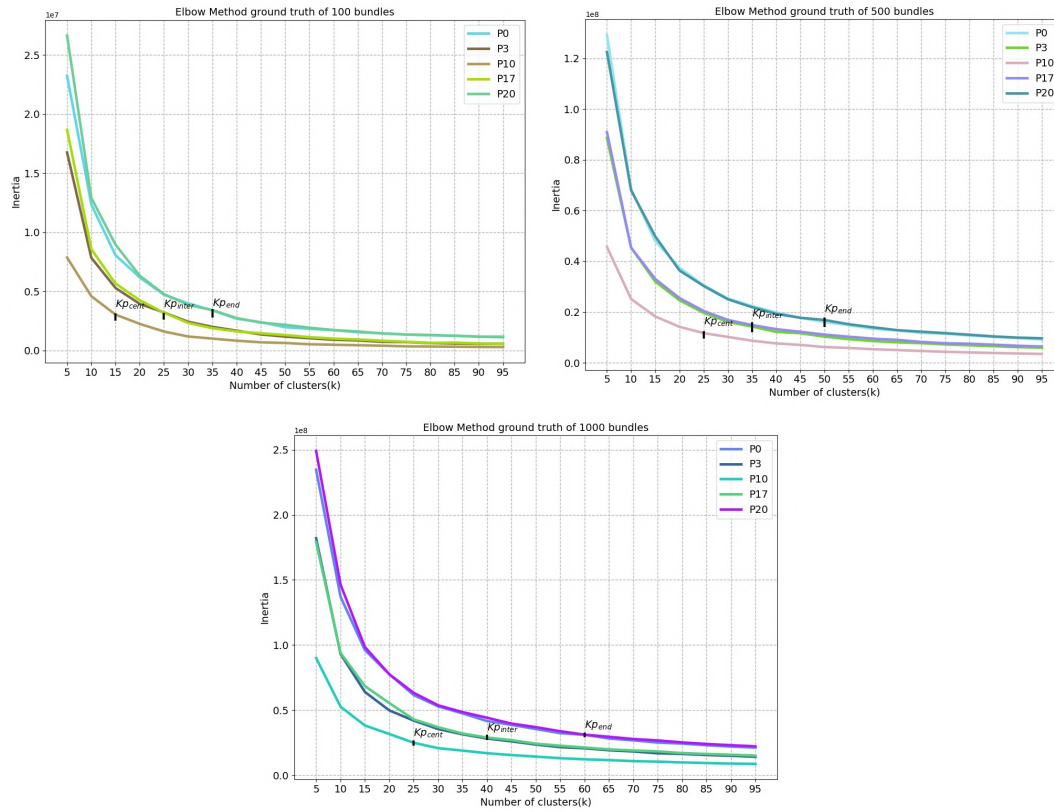


Figura A.1.10: Curvas del método del codo utilizadas para calcular el número de *clusters* en cada punto de fibra (0, 3, 10, 17, 20) para el MiniBatch Kmeans, que corresponde al primer paso del algoritmo FFClust. Los puntos seleccionados para los diferentes conjuntos de datos de referencia son: para 100 fascículos $Kp_{end} = 35$, $Kp_{inter} = 25$, $Kp_{cent} = 15$, 500 fascículos: $Kp_{end} = 50$, $Kp_{inter} = 35$, $Kp_{cent} = 25$, y 1000 fascículos: $Kp_{end} = 60$, $Kp_{inter} = 40$, $Kp_{cent} = 25$.

A.1.5. Métricas para la evaluación de los algoritmos de clustering Quickbundles, FFClust y QBX.

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
True Positives	51.00	79.00	77.00	57.00	156.00	242.00	179.00	71.00	293.00	420.00	276.00	84.00
False Negatives	49.00	21.00	23.00	43.00	344.00	258.00	321.00	429.00	707.00	580.00	724.00	916.00
False Positives	116.00	31.00	14.00	19.00	582.00	226.00	149.00	128.00	1120.00	464.00	296.00	199.00
Clustering-wise Sensitivity	0.78	0.94	0.97	0.98	0.74	0.89	0.94	0.93	0.72	0.86	0.89	0.90
Positive Predictive Value	0.98	0.96	0.90	0.78	0.89	0.83	0.70	0.47	0.85	0.77	0.60	0.34

Tabla A.1.2: Valores de las métricas True Positives (TP), False Negatives (FN), False Positives (FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV). Estas métricas se utilizaron para obtener las métricas principales: Precision, Recall y Geometric Accuracy (Acc), necesarias para evaluar el rendimiento del algoritmo QB en los tres conjuntos de datos de tractografía simulada. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
True Positives	76.00	79.00	83.00	78.00	133.00	177.00	205.00	172.00	164.00	252.00	316.00	270.00
False Negatives	24.00	21.00	17.00	22.00	367.00	323.00	295.00	328.00	836.00	748.00	684.00	730.00
False Positives	100.00	76.00	43.00	39.00	1060.00	801.00	592.00	516.00	2558.00	1919.00	1403.00	1257.00
Clustering-wise Sensitivity	0.90	0.91	0.94	0.94	0.70	0.76	0.79	0.81	0.64	0.70	0.75	0.76
Positive Predictive Value	0.97	0.97	0.96	0.92	0.89	0.88	0.86	0.81	0.85	0.83	0.82	0.76

Tabla A.1.3: Valores de las métricas True Positives (TP), False Negatives (FN), False Positives (FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV). Estas métricas se utilizaron para obtener las métricas principales: Precision, Recall y Geometric Accuracy (Acc), necesarias para evaluar el rendimiento del algoritmo FFClust en los tres conjuntos de datos de tractografía simulada. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.

Original clusters	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
Algorithm clusters	218	146	117	76	1210	761	486	199	2754	1726	943	283
Accuracy	0.86	0.93	0.93	0.87	0.78	0.84	0.80	0.66	0.76	0.79	0.73	0.56
Precision	0.23	0.55	0.70	0.75	0.12	0.30	0.37	0.36	0.09	0.21	0.27	0.30
Recall	0.47	0.74	0.77	0.57	0.25	0.41	0.34	0.14	0.23	0.33	0.24	0.08
F-Measure	0.31	0.63	0.74	0.65	0.16	0.35	0.36	0.20	0.13	0.26	0.25	0.13
Maximum Matching Ratio	0.46	0.72	0.76	0.56	0.24	0.40	0.33	0.14	0.21	0.31	0.23	0.08

Tabla A.1.4: Valores de las métricas para evaluar el rendimiento del algoritmo QBX en los tres conjuntos de datos de tractografía simulada (100, 500 y 1000 fascículos). En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.

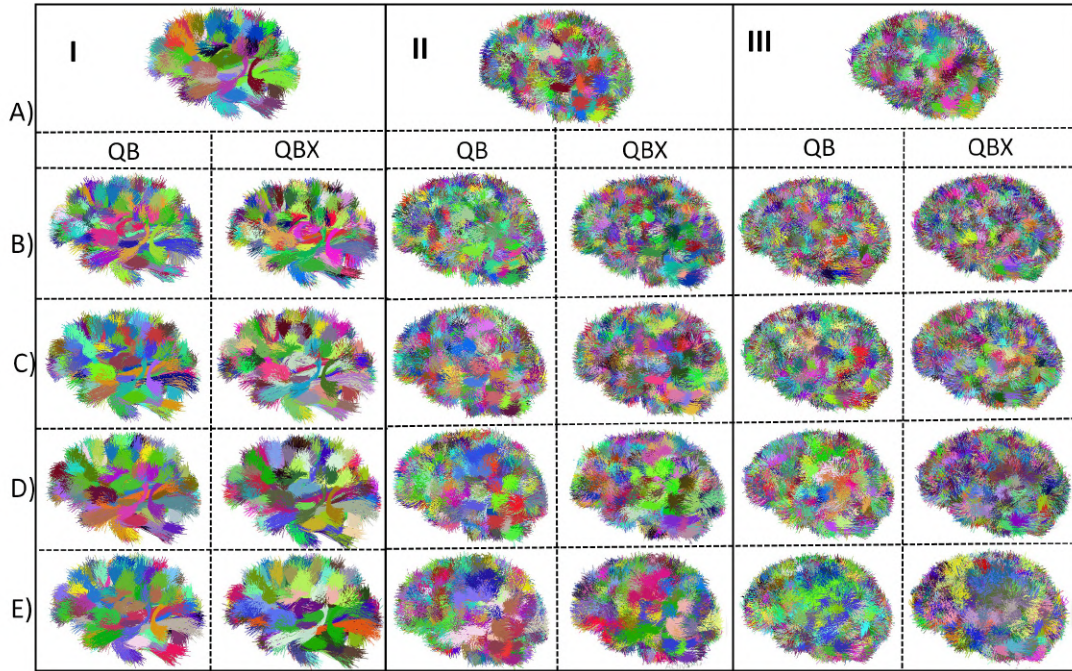


Figura A.1.11: Resultados del *clustering* de fibras obtenidos con los algoritmos QB y QBX para conjuntos de datos de tractografía simulada con 100 fascículos (I), 500 fascículos (II) y 1000 fascículos (III): A) Fascículos simulados originales. *Clusters* obtenidos por QB y QBX para diferentes umbrales: B) 10 mm, C) 12 mm, D) 15 mm, E) 20 mm

Original clusters	100			500				1000				
	10	12	15	20	10	12	15	20	10	12	15	20
Thresholds	46.33	76.00	74.67	47.67	156.00	223.33	161.67	55.00	297.67	381.00	227.00	62.33
True Positives	± 6.43	± 2.83	± 4.04	± 4.04	± 8.19	± 2.52	± 6.11	± 4.58	± 9.71	± 13.11	± 7.94	± 3.51
False Negatives	53.67 ± 6.43	24.00 ± 2.83	25.33 ± 4.04	52.33 ± 4.04	344.00 ± 8.19	276.67 ± 2.52	338.33 ± 6.11	445.00 ± 4.58	702.33 ± 9.71	619.00 ± 13.11	773.00 ± 7.94	937.67 ± 3.51
False Positives	120.67 ± 10.26	38.00 ± 1.41	15.00 ± 3.61	23.00 ± 2.65	582.67 ± 12.58	241.33 ± 2.89	149.33 ± 6.35	127.00 ± 2.00	1113.33 ± 21.94	480.33 ± 20.55	301.00 ± 8.54	200.00 ± 1.00
Clustering-wise Sensitivity	0.77 ± 0.03	0.93 ± 0.00	0.98 ± 0.00	0.98 ± 0.00	0.72 ± 0.01	0.88 ± 0.01	0.93 ± 0.01	0.94 ± 0.01	0.72 ± 0.01	0.85 ± 0.00	0.89 ± 0.01	0.90 ± 0.00
Positive Predictive Value	0.98 ± 0.00	0.95 ± 0.01	0.88 ± 0.01	0.72 ± 0.03	0.87 ± 0.00	0.80 ± 0.00	0.66 ± 0.01	0.43 ± 0.01	0.82 ± 0.00	0.74 ± 0.00	0.55 ± 0.00	0.32 ± 0.01

Tabla A.1.5: Valores de las métricas True Positives (TP), False Negatives (FN), False Positives (FP), Clustering-wise sensitivity (S_n) y Positive Predictive Value (PPV). En azul se destacan las métricas correspondientes al umbral con el mejor rendimiento.

Original clusters Thresholds	100				500				1000			
	10	12	15	20	10	12	15	20	10	12	15	20
True Positives	67.33	76.00	78.67	74.67	136.67	177.67	197.00	176.33	157.67	247.00	309.00	266.00
	±	±	±	±	±	±	±	±	±	±	±	±
False Negatives	2.08	1.73	4.04	2.89	7.23	3.21	9.85	4.73	1.53	11.27	9.17	8.89
	±	±	±	±	±	±	±	±	±	±	±	±
False Positives	32.67	24.00	21.33	25.33	363.33	322.33	303.00	323.67	842.33	753.00	691.00	734.00
	±	±	±	±	±	±	±	±	±	±	±	±
Clustering-wise Sensitivity	121.67	79.67	52.33	46.67	1049.67	786.67	576.33	496.33	2492.67	1903.00	1414.67	1229.00
	±	±	±	±	±	±	±	±	±	±	±	±
Positive Predictive Value	10.97	5.13	7.57	4.16	16.20	2.08	26.01	12.74	7.37	25.24	18.61	8.72
	±	±	±	±	±	±	±	±	±	±	±	±
Positive Predictive Value	0.88	0.91	0.94	0.94	0.71	0.76	0.81	0.82	0.64	0.70	0.75	0.76
	±	±	±	±	±	±	±	±	±	±	±	±
Positive Predictive Value	0.97	0.97	0.97	0.93	0.88	0.87	0.86	0.81	0.84	0.83	0.81	0.76
	±	±	±	±	±	±	±	±	±	±	±	±
Positive Predictive Value	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01
	±	±	±	±	±	±	±	±	±	±	±	±

Tabla A.1.6: Variación de los valores de las métricas True Positives (TP), False Negatives (FN), False Positives (FP), Clustering-wise sensitivity (Sn) y Positive Predictive Value (PPV) al permutar los datos de entrada del algoritmo FFClust y los tres conjuntos de datos. La tabla muestra la media y la desviación estándar de todas las métricas. En azul, se destacan las métricas correspondientes al umbral con el mejor rendimiento.