



CIMAT

Centro de Investigación en Matemáticas, A.C.
Área de Ciencias de la Computación

Nombre de la tesis
*Métodos para Detección de Activaciones en
Campos Aleatorios Discretos*

Que para obtener el grado de:
Maestra en Ciencias con Especialidad en
Computación y Matemáticas Industriales

P R E S E N T A

Ramírez Ramírez Roxana

DIRECTOR DE TESIS:

Dr. José Luis Marroquín Zaleta

Guanajuato, Gto. a 29 de agosto del 2011

Integrantes del Jurado

Presidente: Dr. Johan Jozef Lode Van Horebeek (CIMAT)

Secretario: Dr. Salvador Ruiz Correa (CIMAT)

Vocal y Director de la Tesis: Dr. José Luis Marroquín Zaleta (CIMAT)

Asesor:

Dr. José Luis Marroquín Zaleta.

Sustentante:

Roxana Ramírez Ramírez

Agradecimientos

Agradezco al Dr. José Luis Marroquín Zaleta por los consejos que tan amablemente me proporcionó y por la paciencia con que solucionó mis dudas.

Agradezco a mis padres por el apoyo y confianza que han depositado en las decisiones que he tomado.

Agradezco a Mario porque esta a mi lado siempre que lo necesito, porque me anima a seguir adelante en los momentos más difíciles.

Gracias a mis amigos Aristeo, Faustino, Roberto, Carlos, Francisco, y Martín.

Gracias al Consejo Nacional de Ciencia y Tecnología (CONACyT), por haber subsidiado mis estudios de maestría.

Prefacio

Con el presente trabajo de tesis se concluyen mis estudios de Maestría en Ciencias con Especialidad en Ciencias de la Computación y Matemáticas Industriales, que realicé durante un periodo de dos años que abarcan los meses de Agosto de 2009 a Julio de 2011 en las instalaciones del Centro de Investigación de Matemáticas (CIMAT), el cuál se encuentra ubicado en la ciudad de Guanajuato capital.

Las contribuciones que se alcanzan con esta tesis son:

La presentación de seis nuevos métodos para la detección de activaciones en campos aleatorios discretos. Estos se basan en operadores robustos con soporte local. Dichos operadores son estadísticos de orden, moda, mediana y filtro bilateral Gaussiano.

La estructura del presente trabajo se plantea como sigue:

Capítulo 1. En este capítulo se muestran los conceptos claves para el desarrollo del presente trabajo.

Capítulo 2. Aquí se describen de manera breve algunos de los métodos que se han publicado para detectar activaciones en campos aleatorios discretos.

Capítulo 3. Consiste en la descripción de los métodos basados en familias de operadores para detección de Activaciones en Campos Aleatorios Discretos.

Capítulo 4. Se explican los métodos basados en estimadores robustos para la detección de activaciones en campos aleatorios discretos.

Capítulo 5. Se muestran los experimentos y resultados obtenidos con los métodos que se describen en la presente tesis.

Capítulo 6. Finalmente se muestran las conclusiones y trabajo futuro.

Índice general

1	Introducción	1
1.1	Prueba de Hipótesis	1
1.1.1	Pruebas de Unión-Intersección e Intersección-Unión	2
1.1.2	Probabilidades de Error y la Función Potencia	3
1.1.3	p -Valor	4
1.2	Problema de Comparaciones Múltiples	5
1.3	Procedimientos Basados en Permutación	8
2	Métodos Publicados para Detección de Activaciones en Campos Aleatorios Discretos	9
2.1	Algoritmo Estándar Family-Wise Error Rate (FWER)	9
2.2	Supra Threshold Cluster Size (STCS)	11
2.3	Supra Threshold Cluster Mass (STCM)	12
2.4	Threshold Free Cluster Enhancement (TFCE)	13
3	Métodos Basados en Familias de Operadores para Detección de Activaciones en Campos Aleatorios Discretos	15
3.1	Morphology-Based Hypothesis Testing (MBHT)	16
3.1.1	Propiedades	17
3.1.2	Familias de Elementos Estructurales	22
3.2	Procedimiento Adaptable MBHT	24
3.3	Multi-Threshold Supra Threshold Cluster Size (MTSTCS)	25
3.4	Multi-Threshold Supra Threshold Cluster Mass (MTSTCM)	27
3.5	MBHT-TFCE	29
4	Métodos Basados en Estimadores Robustos para Detección de Activaciones en Campos Aleatorios Discretos	30

4.1	Método Basado en Estadísticos de Orden	31
4.2	Métodos Basados en el Histograma Local Ponderado	34
4.2.1	Método de Cuantil Local	38
4.2.2	Método de Moda Local	40
4.2.3	Método de Filtro Bilateral	40
5	Experimentos y Resultados	46
5.1	Simulación de Campos Gaussianos	46
5.2	Datos Sintéticos fMRI	59
6	Conclusiones	70
	Bibliografía	73

Introducción

En el presente trabajo se plantean nuevos métodos para la detección de activaciones en campos aleatorios discretos, que pueden ser de utilidad para enfrentar el problema de las comparaciones múltiples en neuroimagen. Se hace además una descripción de algunos de los métodos existentes, por ejemplo aquellos que se basan en ideas de conectividad, o que combinan conectividad e intensidad, entre otros.

Para mostrar la capacidad de las diferentes técnicas que aquí se presentan realizaremos ejemplos de sus aplicaciones sobre datos sintéticos que incluyen simulaciones de fMRI, *imágenes de resonancia magnética funcional* (o *functional Magnetic Resonance Imaging*)¹.

También realizaremos comparaciones entre los métodos existentes, algunas extensiones de éstos basados en su funcionamiento básico, y nuestras propuestas, analizando las ventajas y desventajas que se tienen sobre los primeros.

Antes de iniciar es necesario introducir algunos conceptos que sirven como base para el entendimiento de los métodos. Comenzaremos definiendo lo que es una prueba de hipótesis.

1.1. Prueba de Hipótesis

Una hipótesis es una afirmación que hacemos sobre una población. Una prueba de hipótesis es un método de inferencia que, basado en una muestra de la población, determina entre dos hipótesis complementarias cuál se acepta como verdadera. Dichas hipótesis complementarias son llamadas *hipótesis nula* e *hipótesis alternativa*, y se denotan por H_0 y $H_1 = H_0^c$ respectivamente.

Definición 1.1.1. *Una prueba de hipótesis es una regla que especifica:*

1. *Para que valores de la muestra se acepta H_0 como cierta.*

¹fMRI es un tipo especializado de resonancia magnética que mide la respuesta hemodinámica (cambios en el flujo sanguíneo) en relación con la actividad neuronal en el cerebro o la médula espinal de los seres humanos u otros animales.

2. Para que valores de la muestra se rechaza H_0 y se acepta H_1 como verdadera.

Al subconjunto del espacio muestral para el cual se rechaza H_0 se le llama *región de rechazo* o *región crítica*. El complemento de la región de rechazo es llamado *región de aceptación*.

Típicamente, una prueba de hipótesis se especifica a través de una función de la muestra llamada *estadístico de prueba* $T(X_1, \dots, X_n) = T(\mathbf{X})$. Por ejemplo, una prueba puede establecer que H_0 se rechaza si \bar{X} , la media de la muestra, es mayor que 3. En este caso $T(\mathbf{X}) = \bar{X}$ es el estadístico de prueba y la región de rechazo es $\{x_1, \dots, x_n: \bar{x} > 3\}$. Existen varios métodos para elegir el estadístico de prueba y la región de rechazo. Para más información consultar [5, Casella & Berger (2002)].

1.1.1. Pruebas de Unión-Intersección e Intersección-Unión

En algunos casos la hipótesis nula puede expresarse en términos de hipótesis nulas más simples. Hacer esto puede simplificar considerablemente la prueba de hipótesis. A continuación se enuncian dos métodos relacionados a estas ideas.

El *método unión-intersección* puede ser útil cuando la hipótesis nula se puede expresar convenientemente como una intersección, esto es

$$H_0: \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma, \quad (1.1.1)$$

donde Γ es un conjunto de índices arbitrario que puede ser finito o infinito, dependiendo del problema. Suponga que es posible realizar la prueba de hipótesis para cada uno de los problemas $H_{0\gamma}: \theta \in \Theta_\gamma$ contra $H_{1\gamma}: \theta \in \Theta_\gamma^c$. Si la región de rechazo para las pruebas de $H_{0\gamma}$ es $\{\mathbf{x}: T_\gamma(\mathbf{x}) \in R_\gamma\}$, entonces la región de rechazo para la prueba de unión-intersección es

$$\bigcup_{\gamma \in \Gamma} \{\mathbf{x}: T_\gamma(\mathbf{x}) \in R_\gamma\} \quad (1.1.2)$$

La razón es muy simple. De la expresión (1.1.1) sabemos que H_0 se acepta si y sólo si $H_{0\gamma}$ es verdadera para cada γ , entonces si cualquiera de las hipótesis $H_{0\gamma}$ es rechazada, H_0 también será rechazada.

Existen situaciones en las que es posible encontrar una expresión sencilla para la región de rechazo de una prueba de unión-intersección. Por ejemplo, suponga que cada una de las pruebas individuales tiene una región de rechazo de la forma $\{\mathbf{x}: T_\gamma(\mathbf{x}) > c\}$, donde c no depende de γ . Entonces la expresión (1.1.2) para la región de rechazo de la prueba de unión-intersección se puede reescribir como

$$\bigcup_{\gamma \in \Gamma} \{\mathbf{x}: T_\gamma(\mathbf{x}) > c\} = \left\{ \mathbf{x}: \sup_{\gamma \in \Gamma} T_\gamma(\mathbf{x}) > c \right\}. \quad (1.1.3)$$

Luego el estadístico de prueba para H_0 es $T(\mathbf{x}) = \sup_{\gamma \in \Gamma} T_\gamma(\mathbf{x})$.

El *método intersección-uni6n*, puede utilizarse si la hip6tesis nula puede expresarse como una uni6n, es decir

$$H_0: \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma. \quad (1.1.4)$$

Suponga que para cada $\gamma \in \Gamma$, $\{\mathbf{x}: T_\gamma(\mathbf{x}) \in R_\gamma\}$ es la regi6n de rechazo para la prueba $H_{0\gamma}: \theta \in \Theta_\gamma$ contra $H_{1\gamma}: \theta \in \Theta_\gamma^c$. Entonces la regi6n de rechazo para la prueba intersecci6n-uni6n de H_0 contra H_1 es

$$\bigcap_{\gamma \in \Gamma} \{\mathbf{x}: T_\gamma(\mathbf{x}) \in R_\gamma\} \quad (1.1.5)$$

ya que de la expresi6n (1.1.4), H_0 se rechaza si y s6lo si todas las $H_{0\gamma}$ se rechazan. Si la regi6n de rechazo para todas las hip6tesis individuales es de la forma $\{\mathbf{x}: T_\gamma(\mathbf{x}) \geq c\}$ (con c independiente de γ) entonces la regi6n de rechazo para H_0 puede simplificarse como

$$\bigcap_{\gamma \in \Gamma} \{\mathbf{x}: T_\gamma(\mathbf{x}) \geq c\} = \left\{ \mathbf{x}: \inf_{\gamma \in \Gamma} T_\gamma(\mathbf{x}) \geq c \right\}. \quad (1.1.6)$$

Aqu6, el estad6stico de prueba intersecci6n-uni6n es $\inf_{\gamma \in \Gamma} T_\gamma(\mathbf{x})$, y la prueba rechaza H_0 para valores grandes de este estad6stico.

1.1.2. Probabilidades de Error y la Funci6n Potencia

Por lo general, una prueba de hip6tesis es evaluada y comparada a trav6s de su probabilidad de cometer errores. Estas probabilidades de error pueden controlarse en cierta medida. Una prueba de hip6tesis de $H_0: \theta \in \Theta_0$ contra $H_1: \theta \in \Theta_0^c$ puede cometer dos tipos de errores. El *Error de Tipo I* se comete cuando se decide rechazar la hip6tesis nula (H_0) dado que 6sta es verdadera en la poblaci6n. Por otro lado, el *Error de Tipo II* ocurre si la hip6tesis nula (H_0) es aceptada cuando en realidad es falsa sobre la poblaci6n. La tabla 1.1.1 muestra los casos que se pueden presentar.

	H_0 es verdadera	H_1 es verdadera
Se acepta H_0	Sin error (verdadero positivo)	Error Tipo II (falso negativo)
Se acepta H_1	Error Tipo I (falso positivo)	Sin error (verdadero negativo)

Tabla 1.1.1: Tipos de errores en pruebas de hip6tesis.

Dada una prueba de hip6tesis con regi6n de rechazo R , la funci6n de θ , $P_\theta(\mathbf{X} \in R)$, contiene

toda la información sobre la prueba con región de rechazo R , ya que

$$P_{\theta}(\mathbf{X} \in R) = \begin{cases} \text{probabilidad de Error de Tipo } I & \text{si } \theta \in \Theta_0 \\ \text{uno menos la probabilidad de Error de Tipo } II & \text{si } \theta \in \Theta_0^c \end{cases} \quad (1.1.7)$$

Lo anterior nos lleva a la siguiente definición.

Definición 1.1.2. *La función potencia de una prueba de hipótesis con región de rechazo R es la función de θ definida por $\beta(\theta) = P_{\theta}(\mathbf{X} \in R)$.*

La función potencia ideal es 0 para todo $\theta \in \Theta_0$ y 1 para todo $\theta \in \Theta_0^c$. En la práctica esta función ideal no se puede obtener, exceptuando algunas situaciones triviales. Por tanto se debe elegir la prueba de hipótesis cuya estructura de error, obtenida mediante la función potencia, sea la “más conveniente”. Esto es, que el valor de la función potencia este cercano a 1 para la mayoría de $\theta \in \Theta_0^c$ y este cercano a 0 para la mayoría de $\theta \in \Theta_0$.

Por lo general, la función potencia de una prueba dependerá del tamaño n de la muestra. En estos casos, considerar la función potencia podría ayudar a determinar el tamaño de muestra apropiado para un experimento.

En la búsqueda de una buena prueba de hipótesis, es común restringirnos a considerar las pruebas que controlan la probabilidad del Error de Tipo I en un nivel específico, para luego seleccionar del grupo aquella que tenga probabilidad de Error de Tipo II más pequeña posible.

Definición 1.1.3. *Para $0 \leq \alpha \leq 1$, una prueba con función potencia $\beta(\theta)$ es una prueba de nivel α si $\sup_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$.*

Es común que se elija $\alpha = 0.01, 0.05$ y 0.10 . Las pruebas de nivel α tienen probabilidad de Error de Tipo I a lo más α para todo $\theta \in \Theta_0$. Hay que tener en cuenta que al fijar el nivel de la prueba solamente se esta controlando la probabilidad de Error de Tipo I y no la probabilidad de Error de Tipo II . Una buena prueba, dentro de esta clase, es aquella que tiene una probabilidad de Error de Tipo II más pequeña que todas las otra pruebas en la clase.

1.1.3. p -Valor

Una vez que la prueba de hipótesis se ha realizado, es necesario reportar las conclusiones por medio de un método estadístico. Una posibilidad consiste en indicar el nivel α de la prueba usado en la decisión de rechazar H_0 o no rechazar H_0 . Si α es pequeña, la decisión de rechazar H_0 es convincente, pero si α es grande, la decisión de rechazar H_0 es dudosa debido a que la probabilidad de tomar una decisión incorrecta es grande.

Existe otra táctica para reportar los resultados de una prueba de hipótesis que se apoya en el concepto de p -valor. El p -valor se define como la probabilidad de obtener un resultado como

el que realmente se ha obtenido (valor del estadístico calculado), bajo la suposición de que la hipótesis nula es cierta. Si el p -valor ligado al resultado observado es menor o igual que el nivel de significancia fijado, se opta por *no* aceptar la hipótesis nula, ya que el p -valor indica que es más probable que la hipótesis de partida sea incorrecta. No obstante, podría tratarse de una observación atípica, por lo que estaríamos incurriendo en el *error estadístico* de no aceptar la hipótesis nula cuando ésta no debería rechazarse. Este tipo de errores disminuyen incrementando el tamaño de la muestra obtenida, esto reduce el riesgo de que el dato obtenido sea casualmente atípico.

Así, es común decir que valores altos del p -valor no permiten rechazar la hipótesis H_0 . De igual manera, valores bajos del p -valor nos impiden aceptar H_0 . Una ventaja de reportar el resultado de la prueba vía un p -valor es que cada lector puede seleccionar el α que considere apropiado y entonces puede comparar el $p(x)$ reportado en α y saber si estos datos conducen a rechazar o no rechazar H_0 . Además, entre más pequeño sea el p -valor, se tiene mayor certeza para rechazar H_0 .

1.2. Problema de Comparaciones Múltiples

En estadística el problema de comparaciones múltiples (o pruebas múltiples) ocurre cuando se considera un conjunto de inferencias estadísticas de forma simultánea. En pruebas de hipótesis este problema resulta del incremento de las ocurrencias de falsos positivos. Si n comparaciones independientes son realizadas, cada una con un nivel de significancia α_0 , entonces el nivel de significancia del experimento completo está dado por

$$\alpha = 1 - (1 - \alpha_0)^n. \quad (1.2.1)$$

Aquí α representa la probabilidad de cometer uno o más errores de tipo I en las n pruebas, y suele recibir el nombre de *Family-Wise Error Rate* (FWER). Es importante notar que, a menos de que las pruebas sean perfectamente dependientes, α se incrementa a medida que aumenta el número de comparaciones n . Para atacar el problema de comparaciones múltiples es necesario corregir el nivel de significancia con el fin de controlar acertadamente los errores de Tipo I .

Los datos obtenidos con técnicas de *neuroimagen funcional*² representan un problema de comparaciones múltiples. En este caso, una prueba sobre el voxel x corresponde a hallar un umbral de rechazo. Un voxel se considera activo si el estadístico de prueba $t(x)$ alcanza un umbral u dado. La hipótesis nula H_{0x} establece que no hay activación en dicho voxel. Al considerar cada voxel separadamente, hasta cientos de miles de pruebas se llevan a cabo.

²Las Técnicas de neuroimagen funcional, desarrolladas en la década de los 80, permiten interpretar en imágenes tomográficas los cambios cerebrales que se producen mientras el sujeto está realizando una actividad cognitiva. Éstas incluyen la Tomografía por Emisión de Positrones (PET), la Tomografía Computarizada por Emisión de Fotones Simples (SPECT) y la Resonancia Magnética Funcional (fMRI).

Una prueba de nivel α sobre el voxel x satisface $P[t(x) \geq u | H_{0x} \text{ cierta}] \leq \alpha$. Entonces la dificultad a la que uno se enfrenta cuando tiene un problema de pruebas múltiples es encontrar un umbral u que controle en alguna medida los falsos positivos a través de toda la imagen.

Se han propuesto diversas soluciones para controlar los falsos positivos en experimentos de neuroimagen, en la mayoría de éstas se usa el *family-wise error rate* como la medida de interés (por ejemplo el algoritmo estándar FWER, el cuál se describe en la sección 2.1). Para información más detallada sobre como controlar the family-wise error rate en neuroimagen funcional, véase [22, Nichols and Hayasaka (2003)].

Algunos estadistas sugieren usar el procedimiento llamado *Corrección de Bonferroni* o *Sidák* [1, Abdi (2007)] por su simplicidad además de que es muy intuitivo. La idea es la siguiente: si se están probando n hipótesis dependientes o independientes en un conjunto de datos, para evaluar las comparaciones entre éstas con un nivel de significancia α , cada una de las comparaciones debe realizarse al nivel α/n . Desafortunadamente este planteamiento suele incrementar el error de Tipo II y por tanto reduce la potencia estadística.

En la actualidad la manera más usada para abordar el problema se apoya en la teoría de extremos de campos aleatorios Gaussianos [34, Worsley et al., (1992)] y se encuentra implementado en la mayoría de los paquetes usados para el análisis en neuroimagen. Este enfoque es muy elegante y tiene fundamentos teóricos sólidos; sin embargo, hace algunos supuestos fuertes sobre la distribución y la correlación de los datos, los cuales pueden ser desconocidos en la práctica ([35, Zarahn et al., (1997)], [26, Purdon and Weisskoff (1998)]); otro de sus inconvenientes es que requiere que los datos sean suavizados antes de procesarlos, y ésto reduce la resolución espacial de los resultados. Además, este método es ligeramente menos conservador que la corrección de Bonferroni, en particular para datos con suavizado sutil. Debido a esto, los umbrales estadísticos usados para reportar activaciones significativas en estudios de neuroimagen funcional varían ampliamente, lo que hace difícil comparar resultados entre estudios [30, Sergerie et al. (2008)].

Otros enfoques que también se han desarrollado con la finalidad de aplicarse al problema de las comparaciones múltiples en neuroimagen son: *False Discovery Rate* (FDR) [12, Genovese et al., (2002)], [3, Benjamini and Hochberg, (1995)], el cual se vuelve cada vez más popular; y la tasa de error por comparación [21, Nichols, (2002)], por mencionar algunos.

Es importante hacer notar que en todos estos métodos la decisión de declarar como activo o no activo un voxel en particular, solo depende de la magnitud de la señal en ese voxel independientemente de la magnitud que tengan sus vecinos (aunque los umbrales de decisión se determinan tomando en cuenta la dependencia espacial bajo la hipótesis nula, y el dato actual de voxeles vecinos también se considera hasta un cierto punto, debido al suavizado que se le aplica a la imagen). Sin embargo, es muy raro que las activaciones se presenten en voxeles aislados, es más común que se trate de una región de voxeles contiguos (que nosotros llamaremos cluster) debido a la correlación espacial que existe en la señal.

Tomando lo anterior en cuenta, han surgido otras soluciones al problema de las comparaciones múltiples. Algunos de estos métodos usan el concepto de conectividad en la rejilla de voxeles para definir agrupaciones como activas o no activas, dónde el tamaño de éstas es usado como una métrica [10, Friston et al., (1994b)]. Estos métodos tienden a ser más susceptibles a detectar activaciones débiles con extensión espacial amplia, mientras que las pruebas que se sustentan únicamente en los voxeles son más aptas para reconocer activaciones localizadas de magnitud grande. Este enfoque también tiene algunas limitaciones; la principal de ellas es que la elección del valor del umbral es crítica, y no cuenta con fundamentos fuertes para determinarlo. Otra de sus limitantes es que la noción de conectividad en una rejilla no captura bien el concepto intuitivo de un “cluster compacto contiguo”, ya que conjuntos de voxeles conectados pueden formar estructuras grandes porosas (como un árbol en imágenes de ruido), lo que prodría ser clasificado erróneamente por el método como una región activa (ver figura 5.1.11). Otro problema que surge al aplicar una prueba al tamaño del cluster es que regiones pequeñas de intensidad alta podrían no ser detectadas, aunque ellas correspondan a activaciones significativas.

Se han planteado algunas soluciones a este último problema: en [25, Poline et al. (1997)] se propone una prueba que combina la intensidad del voxel y la extensión del cluster basada en Teoría de Campos Aleatorios, aunque hacen supuestos fuertes sobre la distribución que pueden no conocerse en la práctica. Otros métodos que combinan la intensidad en el voxel y el tamaño del cluster están basados en la idea de calcular la distribución empírica de la masa³ del cluster ([4, Bullmore et al., (1999)], [36, Zhang et al., (2009)]), pero el estadístico basado en esta distribución puede fallar al detectar activaciones en la señal cuando sólo la intensidad o extensión espacial es significativa (parece no haber dificultad cuando ambas son significativas).

Esta prueba y otras funciones que combinan estadísticos de intensidad y extensión pueden ser implementadas usando métodos basados en permutaciones [13, Hayasaka and Nichols, (2004)]. En consecuencia son robustas con respecto a suposiciones precisas de la distribución, pero además del problema de la selección del umbral, estos procedimientos pueden aun fallar para detectar pequeñas agrupaciones (mayores a un voxel) con un nivel de activación moderado. Recientemente se ha presentado un método simple y elegante que combina la intensidad y el tamaño del cluster en [32, Smith and Nichols, (2009)]; en este enfoque un estadístico de prueba que combina tamaños de cluster, que se encuentran usando un conjunto de umbrales de diferentes intensidades, es usado en un ajuste no-paramétrico evitando así el problema de la selección del umbral. Este método se revisa a detalle en la sección 2.4.

Otros métodos son extensiones del enfoque *False Discovery Rate* (FDR), así que pueden ser aplicados ya sea a datos pre-agrupados ([14, Heller et al., (2006)], [2, Benjamini and Heller, (2007)]), o a datos crudos, pero a nivel regional [18, Langer et al., (2007)]. Estos métodos usan modelos paramétricos para calcular los p -valores (ya sea para agrupaciones o para voxeles

³La masa es la suma de las intensidades de todos los voxeles que pertenecen al cluster.

individuales) los cuales se usan para detectar regiones activas en las que el FDR esperado se controla a un nivel deseado. Una idea similar se usa en [6, Chumbley and Friston (2009)] y [7, Chumbley et al., (2009)] donde las distribuciones nulas teóricas para el volumen del cluster y para picos (máximos locales) en el interior de un cluster son derivadas por campos Gaussianos suaves usando Teoría de Campos Aleatorios.

1.3. Procedimientos Basados en Permutación

Para estimar la distribución del estadístico de prueba bajo H_0 es necesario producir muchas muestras bajo la hipótesis de partida. No obstante, es común que en aplicaciones prácticas se carezca de un modelo teórico M_0 fiable que nos permita generar campos muestrales del estadístico de prueba que satisfagan H_0 , y por tanto, uno debe recurrir a métodos basados en permutaciones. Al permutar aleatoriamente las observaciones se pueden generar nuevos conjuntos muestrales.

Los procedimientos basados en permutaciones son una manera simple de producir muestras de la distribución de cualquier estadístico de prueba siempre que las observaciones satisfagan ciertos requerimientos. Por lo general se consideran los siguientes dos ([19, Lehmann and Romano, (2005)], [15, Holmes et al., (1996)]):

1. Supuestos débiles sobre la distribución que aseguren algún tipo de intercambiabilidad bajo la hipótesis H_0 . Esto es, si se efectuaran intercambios en la muestra, éstos no tendrían efecto alguno sobre el resultado. Por ejemplo, intercambiabilidad de sujetos, de residuos, de modelos auto-regresivos de ajuste, etc.
 2. Aleatorización experimental. Por ejemplo, la aleatorización de estímulos como la que sugieren en [27, Raz et al., (2003)]. En este caso, una prueba de permutación se puede construir permutando las etiquetas de estímulos en la misma forma en que éstas se asignaron por azar en el experimento. Los ensayos realizados para fMRI en la sección 5.2 son ejemplo de esto. Ver secciones 5.8-5.15 de [19, Lehmann and Romano, (2005)].
-

Métodos Publicados para Detección de Activaciones en Campos Aleatorios Discretos

En este capítulo se dará una descripción muy breve de algunos de los métodos que se han publicado para detectar activaciones en campos aleatorios discretos.

Como ya se mencionó en la sección 1.2, analizar las imágenes obtenidas con técnicas de neuroimagen se convierte en un problema de múltiples comparaciones. A continuación definiremos la hipótesis nula sobre el campo de medidas de interés de la siguiente forma:

$$H_0 = \bigcap_{\mathbf{x} \in L} H_{0\mathbf{x}}, \quad (2.0.1)$$

donde $H_{0\mathbf{x}}$ es una hipótesis nula marginal sobre la distribución de probabilidad de las mediciones en el voxel $x \in L$ (L es la rejilla de voxels). También considere un estadístico de prueba marginal adecuado $t(\mathbf{x})$ para el problema de $H_{0\mathbf{x}}$ contra $H_{0\mathbf{x}}^c$ en cada voxel, con lo que definimos al campo de estadísticos de pruebas como $\{t(\mathbf{x}) : \mathbf{x} \in L\}$. Por ejemplo, en el caso de imágenes funcionales, la hipótesis alternativa $H_{1\mathbf{x}} = H_{0\mathbf{x}}^c$ puede representar la hipótesis de que el voxel \mathbf{x} está activado bajo alguna condición experimental, al comparar con una condición de referencia; y $t(\mathbf{x})$ puede ser un estadístico t o F para probar esta hipótesis. En general, a los voxels $\mathbf{x} \in L$ donde $H_{0\mathbf{x}}$ es falsa los llamaremos *voxels activos*, y a la colección de éstos como *Región de Activación* (RA), la cuál puede consistir de diversas agrupaciones, y simplemente la denotaremos por A .

Con la notación anterior ya podemos dar pie a la explicación de los métodos.

2.1. Algoritmo Estándar Family-Wise Error Rate (FWER)

Suponga que contamos con un modelo M_0 para generar campos t del estadístico de prueba bajo la hipótesis nula H_0 (por ejemplo, M_0 podría ser un procedimiento basado en permutaciones).

La manera estándar (no-paramétrica) de controlar el FWER comienza con la determinación de los umbrales apropiados basados en la distribución del máximo. Esto se puede hacer siguiendo el algoritmo 2.1.1:

Algorithm 2.1.1: FWER

Input: El modelo estadístico M_0 para generar campos bajo H_0 y un nivel de significancia α .

Output: el umbral θ_0^* .

- 1 Generar muestras de campos t usando M_0 .
 - 2 Para cada muestra t calcular la variable aleatoria $\theta_0 = \max_{\mathbf{x} \in L} t(\mathbf{x})$.
 - 3 Encontrar la distribución empírica $F_{\theta_0}(\cdot)$ de los valores θ_0 .
 - 4 Para un nivel de significancia α dado, encontrar θ_0^* tal que $F_{\theta_0}(\theta_0^*) = 1 - \alpha$.
 - 5 **return** θ_0^* .
-

Luego la RA estimada \hat{A}_0 esta dada por

$$\hat{A}_0 = u_{\theta_0^*} [t], \tag{2.1.1}$$

donde u_θ es el operador umbral, esto es, $u_\theta [t] (\mathbf{x}) = 1$ si $t(\mathbf{x}) > \theta$ y cero en otro caso.

Si luego aplicamos el umbral θ_0^* a un campo de estadísticos de prueba t que no contiene activaciones, esto es, si la hipótesis nula se mantiene, podemos asegurar con un nivel de confianza de $1 - \alpha$ que no se encontrarán activaciones falsas (voxeles con valor de 1) en la imagen binaria $u_{\theta_0^*} [t]$. Esta propiedad es llamada *Control Débil del Error* (WEC, Weak Error Control)[19, Lehmann and Romano, (2005)].

Además, para cualquier subconjunto $S \subseteq L$ tal que $S \cap A = \emptyset$, la probabilidad de rechazar la hipótesis $H_{0\mathbf{x}}$ para cualquier $\mathbf{x} \in S$ (esto es, la probabilidad de que $u_{\theta_0^*} [t] = 1$ para algún $\mathbf{x} \in S$) también esta acotada por α , es decir que FWER tiene *Control Fuerte del Error* (SEC, Strong Error Control). Esta propiedad se pierde si se aplica un suavizado a los datos. Notar que θ_0^* es mayor que el umbral que rechaza la hipótesis nula marginal $H_{0\mathbf{x}}$ para un voxel fijo \mathbf{x} con nivel de significancia α . Esto se debe a que el cálculo del umbral involucra $|L|$ pruebas de hipótesis (donde $|L|$ denota la cardinalidad de L), por lo que esperaríamos en promedio $\alpha|L|$ activaciones falsas.

Un método con control débil solamente da certeza de que H_0 es falsa, pero no permite la localización de voxeles significativos independientes. Debido a esto, algunas veces a las pruebas que sólo tienen control débil se les llama pruebas “generales”. En cambio un método que cuenta con control fuerte nos permite rechazar las hipótesis marginales $H_{0\mathbf{x}}$ mientras se controla el FWER en todos los voxeles no significativos.

2.2. Supra Threshold Cluster Size (STCS)

Este procedimiento se plantea en [10, Friston et al., (1994b)] y forma parte de las técnicas basadas en estadísticos de *Supra-Threshold Connected Cluster* (STCC, grupos conectados que están por encima de un umbral). Básicamente consiste en calcular la distribución del máximo del tamaño del clúster de voxeles conectados. El clúster esta conformado por los voxeles conectados de la imagen que superan un determinado umbral; este último a menudo se especifica por medio de una heurística adecuada.

Esta técnica surge debido a que, una región de activación se caracteriza no solo por los voxeles que alcanzan un umbral de intensidad determinado, sino también por su topografía espacial, por ejemplo: su forma, cómo se relacionan los voxeles del clúster espacialmente entre sí, y el tamaño de las agrupaciones. Por lo anterior es que en [10, Friston et al., (1994b)] se centran en el tamaño (área o volumen del clúster). Además toman en cuenta que una región espacialmente limitada frecuentemente se considera “menos significativa” que una región de mayor amplitud, esto en base a los estudios empíricos realizados por [24, Poline and Mazoyer, (1993)] y [28, Roland et al., (1993)].

Este enfoque se basa en preguntar: ¿Cuál es el valor crítico para el tamaño máximo de una región?. Y para contestar a ello desde una perspectiva estadística, los autores hacen ver que hay tres cosas que interesan del campo de estadísticos de prueba:

1. El número N de voxeles \mathbf{x} tales que su intensidad sobrepasa un umbral, $t(\mathbf{x}) > u$.
2. El número m de regiones activadas (subconjuntos conectados de voxeles).
3. El número n de voxeles en cada clúster.

Cada uno de estos números tiene su propia función de densidad de probabilidad $P(N = x)$, $P(m = x)$ y $P(n = x)$ (notar que en este contexto $P(n = x)$ esta estrictamente condicionada a que exista al menos una región de activación, $P(n = x | m \geq 1)$). Según los autores estas funciones de probabilidad proveen de una completa caracterización de los campos de estadísticos de prueba.

En esta técnica la probabilidad de interés es: *la probabilidad de obtener al menos una activación con θ_0^* voxeles o más bajo la hipótesis nula*. Esto es equivalente a *la probabilidad de que la región más grande tenga θ_0^* o más voxeles* $P(n_{\text{máx}} \geq \theta_0^* | H_0)$, donde $n_{\text{máx}}$ es el número de voxeles en la región más grande. A continuación se expone el algoritmo 2.2.1 que se basa en la idea principal de este procedimiento.

Algorithm 2.2.1: STCS

Input: El modelo estadístico M_0 para generar campos bajo H_0 , un nivel de significancia α y un umbral de intensidad apropiado u_0 .

Output: el valor crítico de área o volumen θ_0^* .

- 1 Generar muestras de campos t usando M_0 .
 - 2 Para cada muestra t aplicar el operador umbral $u_0[t]$.
 - 3 Agrupar los voxeles 8–conectados \mathbf{x} que hayan pasado el umbral en clústers.
 - 4 Para cada $\mathbf{x} \in L$ calcular el área $a_{u_0}(\mathbf{x})$ del clúster al que pertenece \mathbf{x} .
 - 5 Calcular la variable aleatoria $\theta_0 = \max_{\mathbf{x} \in L} a_{u_0}(\mathbf{x})$.
 - 6 Encontrar la distribución empírica $F_{\theta_0}(\cdot)$ de los valores θ_0 .
 - 7 Para un nivel de significancia α dado, encontrar θ_0^* tal que $F_{\theta_0}(\theta_0^*) = 1 - \alpha$.
 - 8 **return** θ_0^* .
-

2.3. Supra Threshold Cluster Mass (STCM)

Esta técnica se propone en [4, Bullmore et al., (1999)] y al igual que STCS también forma parte de las técnicas basadas en estadísticos de STCC. La idea fundamental de este método radica en calcular la distribución del máximo de la *masa del clúster* de voxeles conectados. Al igual que STCS los clusters son conjuntos de voxeles 8–conectados que exceden un umbral u , determinado a través de una heurística adecuada.

Para calcular la *masa* de un clúster C es necesario sumar las diferencias entre la intensidad de cada voxel $x \in C$ y el umbral u como se muestra en la ecuación (2.3.1) del algoritmo 2.3.1 que detalla el procedimiento de STCM.

El desempeño de los métodos basados en STCC depende fuertemente de la selección de un umbral de intensidad u conveniente; sin embargo, estas técnicas no nos proporcionan un criterio con bases teóricas para elegir dicho umbral, sino que su elección se hace de manera arbitraria.

Algorithm 2.3.1: STCM

Input: El modelo estadístico M_0 para generar campos bajo H_0 , un nivel de significancia α y un umbral de intensidad apropiado u_0 .

Output: el valor crítico de masa θ_0^* .

- 1 Generar muestras de campos t usando M_0 .
- 2 Para cada muestra t aplicar el operador umbral $u_0 [t]$.
- 3 Agrupar los voxeles 8-conectados \mathbf{x} que hayan pasado el umbral en clústers.
- 4 Para cada $\mathbf{x} \in L$ calcular la masa $m_{u_0}(\mathbf{x})$ del clúster C al que pertenece \mathbf{x} con

$$m_{u_0}(\mathbf{x}) = \sum_{\mathbf{x} \in C} h(\mathbf{x}) \quad y \quad h(\mathbf{x}) = \begin{cases} t(\mathbf{x}) - u_0, & \text{si } t(\mathbf{x}) > u_0; \\ 0, & \text{si } t(\mathbf{x}) \leq u_0. \end{cases} \quad (2.3.1)$$

- 5 Calcular la variable aleatoria $\theta_0 = \max_{\mathbf{x} \in L} m_{u_0}(\mathbf{x})$.
 - 6 Encontrar la distribución empírica $F_{\theta_0}(\cdot)$ de los valores θ_0 .
 - 7 Para un nivel de significancia α dado, encontrar θ_0^* tal que $F_{\theta_0}(\theta_0^*) = 1 - \alpha$.
 - 8 **return** θ_0^* .
-

2.4. Threshold Free Cluster Enhancement (TFCE)

Este método lo formularon [32, Smith and Nichols, (2009)] para enfrentar el problema que surge de seleccionar adecuadamente el umbral de intensidad en las técnicas que se basan en la detección de clusters que superan un umbral. Este procedimiento trata de mantener los beneficios de la sensibilidad que dichas técnicas ofrecen. El método toma un campo de estadísticos de prueba (imagen) como entrada y como salida produce una imagen en la que los valores de los voxeles representan la cantidad de soporte espacial local de un clúster.

El enfoque *TFCE* tiene como objetivo mejorar las áreas de la señal que muestran alguna contigüidad espacial sin depender fuertemente de las agrupaciones obtenidas al aplicar el operador de umbral. La imagen pasa a través de un algoritmo que, por decirlo de alguna manera, mejora la intensidad dentro de regiones agrupadas respecto al resto de la imagen (lo que se asume como ruido). Por tanto la imagen de salida no está intrínsecamente agrupada por la aplicación de un umbral, pero se espera que después de aplicar TFCE mejore significativamente la diferencia entre ruido y la RA.

La idea principal con que trabaja el algoritmo es asignar a cada voxel un valor nuevo. Este nuevo valor está dado por la función (2.4.1).

$$TFCE(\mathbf{x}) = \sum_{j=1}^N e(\mathbf{x}, h_j)^E (h_j - h_0)^H \quad (2.4.1)$$

donde $\{h_1, \dots, h_N\}$ es una secuencia creciente de umbrales de intensidad; h_0 es un umbral de referencia, en el artículo [32, Smith and Nichols, (2009)] fijan $h_0 = 0$ y, $e(\mathbf{x}, h_j)$ es el tamaño del clúster conectado en $u_{h_j}[t]$ al cual pertenece el voxel \mathbf{x} . Los valores recomendados para los exponentes son $E = 0.5$ y $H = 0.2$. El valor crítico para el estadístico $TFCE(\mathbf{x})$ dado por la ecuación (2.4.1) también puede obtenerse de manera no paramétrica usando procedimientos de permutación.

Métodos Basados en Familias de Operadores para Detección de Activaciones en Campos Aleatorios Discretos

Algunos de los métodos que presentaremos en este capítulo utilizan los siguientes principios de morfología matemática: erosión y dilatación. Estos principios nos permiten detectar regiones de activación en campos aleatorios con niveles de activación moderados y de extensión espacial relativamente grande. Además de tener presente la magnitud de la señal en cada voxel, estos procesos también consideran la magnitud de sus vecinos. Estos métodos fueron desarrollados en [16, Marroquín et al., (2011)]. A continuación se enuncian los conceptos de morfología matemática que utilizaremos.

La morfología matemática (para más detalles ver [31, Serra, (1984)] y [33, Soille, (2004)]) es una teoría originalmente desarrollada para analizar imágenes binarias con respecto a forma y tamaño. Algunos de los conceptos básicos de esta teoría que nos serán de utilidad son:

- *Elemento Estructural* (SE, Structuring Element) W , que consiste de un conjunto de voxeles que satisfacen una restricción geométrica, por ejemplo, una esfera de radio r centrada en el origen.
- El operador *Erosión* \mathbf{E}_W , con respecto a un SE W .
- El operador *Dilatación* \mathbf{D}_W , asociado a un SE W .

La traslación de W a un voxel dado \mathbf{x} , denotado por $W_{\mathbf{x}}$, se define como el conjunto de voxeles alrededor de \mathbf{x} que satisfacen la restricción correspondiente. Por ejemplo si el SE es $W = \{\mathbf{x}: \|\mathbf{x}\| \leq r\}$ entonces $W_{\mathbf{x}} = \{\mathbf{y}: \|\mathbf{x} - \mathbf{y}\| \leq r\}$. Los operadores de erosión y dilatación se definen como:

$$\mathbf{E}_W[B](\mathbf{x}) = \min_{\mathbf{y} \in W_{\mathbf{x}}} B(\mathbf{y}), \quad (3.0.1)$$

$$\mathbf{D}_W[B](\mathbf{x}) = \max_{\mathbf{y} \in W_{\mathbf{x}}} B(\mathbf{y}), \quad (3.0.2)$$

donde B es un campo binario, esto es, $B(\mathbf{x}) \in \{0, 1\}$ para todo $\mathbf{x} \in L$, y L es la rejilla de voxeles.

Los siguientes métodos amplían el uso de operadores morfológicos al análisis estadístico de las regiones de activación bajo una hipótesis nula en particular.

3.1. Morphology-Based Hypothesis Testing (MBHT)

Este método es una extensión de la técnica FWER y permite un control apropiado de falsos positivos sin tener que ajustar algún parámetro adicional además del nivel de significancia. También se puede adaptar fácilmente a procedimientos basados en permutación, tomando en cuenta las restricciones usuales, y por tanto es una metodología no paramétrica ya que no requiere de suposiciones fuertes sobre la distribución de los datos, ni de la correlación espacio-temporal que existe entre éstos. En este enfoque se consideran las propiedades estadísticas individuales en cada voxel, así como las propiedades estadísticas de sus vecinos.

La idea básica es la siguiente: en lugar de determinar si el valor del estadístico de prueba en cada voxel \mathbf{x} esta sobre un umbral dado (el cuál podría obtenerse como en FWER), uno determina si el campo de prueba sobre un conjunto de voxeles alrededor de \mathbf{x} excede un umbral. Aquí el punto clave es cómo seleccionar este umbral usando criterios estadísticos rigurosos.

Suponiendo que contamos con un modelo M_0 para generar campos t del estadístico de prueba de interés bajo la hipótesis nula, la manera estándar no-paramétrica de controlar el FWER comienza con la determinación de los umbrales apropiados basados en la distribución del máximo, que se puede realizar mediante el algoritmo 2.1.1 visto en la sección 2.1. En este nuevo enfoque, se explota la correlación de conjuntos de voxeles vecinos en el campo t para mejorar la sensibilidad del procedimiento FWER, y al mismo tiempo se mantiene un control apropiado de la tasa de errores de falsos positivos.

Es importante hacer notar que si la erosión de $u_\theta[t]$ es distinta de cero en el voxel \mathbf{x} , esto es $\mathbf{E}_W[u_\theta[t]](\mathbf{x}) = 1$, entonces el valor de t en todo el conjunto $W_\mathbf{x}$ esta por encima del umbral θ . Por lo tanto, la extensión del criterio FWER a conjuntos $W_\mathbf{x}$ implica obtener un umbral θ_W^* tal que la erosión bajo un SE W del campo binario $u_{\theta_W^*}[t]$ sea idénticamente cero con probabilidad $1 - \alpha$, para los campos generados por M_0 . Si luego aplicamos este umbral al campo de prueba t y encontramos que $\mathbf{E}_W[u_{\theta_W^*}[t]](\mathbf{x}) = 1$, para algún voxel \mathbf{x} , significa que L contiene una región activa no falsa $A \neq \emptyset$ tal que $W_\mathbf{x} \cap A \neq \emptyset$ con probabilidad de error menor o igual a α . El umbral θ_W^* se puede obtener con el algoritmo 3.1.1.

Algorithm 3.1.1: MBHT

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α y un SE W .

Output: el umbral de detección θ_W^* .

- 1 Generar muestras de campos t usando M_0 .
- 2 Para cada muestra t calcular la variable aleatoria

$$\theta_W = \max_{\mathbf{x} \in L} \min_{\mathbf{y} \in W_{\mathbf{x}}} t(\mathbf{y}). \quad (3.1.1)$$

- 3 Encontrar la distribución empírica $F_{\theta_W}(\cdot)$ de los valores θ_W .
 - 4 Para un nivel de significancia α dado, encontrar θ_W^* tal que $F_{\theta_W}(\theta_W^*) = 1 - \alpha$.
 - 5 **return** θ_W^* .
-

Al igual que el caso de FWER, si aplicamos el umbral θ_W^* a un campo generado por M_0 , podemos garantizar con confianza de $1 - \alpha$ que no se encontraran activaciones falsas (voxels con valor igual a 1 en $\mathbf{E}_W[u_{\theta_W^*}[t]](\mathbf{x})$). Por consiguiente, este procedimiento también tiene WEC. Con respecto a SEC, considere la hipótesis nula local $\hat{H}_{0\mathbf{x}}: W_{\mathbf{x}} \cap A = \emptyset$, cuya intersección nos da la misma hipótesis nula global como antes. Para cualquier conjunto $S \subseteq L$ para el cual se cumple $\hat{H}_{0\mathbf{x}}: W_{\mathbf{x}} \cap A = \emptyset$, para todo $\mathbf{x} \in S$, la probabilidad de rechazar una o más hipótesis locales $\hat{H}_{0\mathbf{x}}$ esta acotada por α . Sin embargo, hay que notar que la región donde no se tiene la hipótesis nula local $\hat{H}_{0\mathbf{x}}$ no es la RA A , sino que más bien se trata de la dilatación de la RA $\mathbf{D}_W[A]$, de manera que no se tiene el SEC en un sentido estricto. La forma precisa de control del error que se puede lograr con este proceso se discute más adelante.

El p -valor p localizado en \mathbf{x} asociado con la detección de RA basada en MBHT esta dada por

$$p(\mathbf{x}) = 1 - F_{\theta_W}(\theta_W(\mathbf{x})), \quad (3.1.2)$$

donde

$$\theta_W(\mathbf{x}) = \min_{\mathbf{y} \in W_{\mathbf{x}}} t(\mathbf{y}). \quad (3.1.3)$$

3.1.1. Propiedades

Para dar una idea más intuitiva sobre como se comporta el método MBHT en [16, Marroquín et al., (2011)] se deducen algunas fórmulas para los casos donde esto es posible. En particular se muestran expresiones aproximadas para la distribución F_{θ_W} (y por tanto, para la detección de umbrales), y se analiza la sensibilidad de estos umbrales con respecto al nivel de significancia. Además se muestran expresiones para la potencia y la probabilidad de detecciones falsas para ciertos casos simples.

También se establecen las características principales del procedimiento MBHT en el caso general, a saber, este procedimiento produce mejoras en la potencia de detección mientras mantiene bajo control la detección de detecciones falsas.

Umbrales de Detección

En [16, Marroquín et al., (2011)] calculan la distribución de θ_W cuando el modelo M_0 produce campos sin dependencia espacial. Esto se hace con el objetivo de analizar la sensibilidad de los umbrales de detección con respecto al tamaño del SE W y el nivel de significancia.

Si en el campo t no hay correlación espacial, tenemos lo siguiente:

$$P(\theta_W(\mathbf{x}) > \theta) = \prod_{\mathbf{y} \in W_{\mathbf{x}}} P(t(\mathbf{y}) > \theta) = \Phi^{|W|}(\theta), \quad (3.1.4)$$

donde $\theta_W(\mathbf{x})$ está dado por la ecuación (3.1.3), $|W|$ denota la cardinalidad del SE W , y

$$\Phi(\theta) = P_{H_0}(t(\mathbf{x}) > \theta), \quad (3.1.5)$$

que se asume igual para todo $\mathbf{x} \in L$. Por ejemplo, para ruido Gaussiano con media 0, varianza 1 y $\theta > 0$, Φ es igual a:

$$\begin{aligned} \Phi(\theta) &= \frac{1}{\sqrt{2\pi}} \int_{\theta}^{\infty} e^{-\mathbf{x}^2/2} d\mathbf{x} \\ &= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\theta} e^{-\mathbf{x}^2/2} d\mathbf{x} \\ &= 1 - \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^0 e^{-\mathbf{x}^2/2} d\mathbf{x} + \int_0^{\theta} e^{-\mathbf{x}^2/2} d\mathbf{x} \right] \\ &= 1 - \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_0^{\theta} e^{-\mathbf{x}^2/2} d\mathbf{x} \\ &= \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{\theta} \frac{1}{\sqrt{2}} e^{-\mathbf{x}^2/2} d\mathbf{x} \\ &= \frac{1}{2} - \frac{1}{\sqrt{\pi}} \int_0^{\theta/\sqrt{2}} e^{-\mathbf{u}^2} d\mathbf{u} \quad \text{con } \mathbf{u} = \frac{1}{\sqrt{2}}\mathbf{x} \\ &= \frac{1}{2} - \frac{1}{2} \frac{2}{\sqrt{\pi}} \int_0^{\theta/\sqrt{2}} e^{-\mathbf{u}^2} d\mathbf{u} \\ &= \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{\theta}{\sqrt{2}} \right) \right), \end{aligned} \quad (3.1.6)$$

donde $\operatorname{erf}(\cdot)$ es la función de error estándar¹. La distribución acumulada de θ_W es

$$F_{\theta_W}(\theta) = P(\theta_W \leq \theta) = P \left(\max_{\mathbf{x} \in L} \theta_W(\mathbf{x}) \leq \theta \right), \quad (3.1.7)$$

¹Si los resultados de una serie de mediciones se describen a través de una distribución normal con desviación estándar σ y media $\mu = 0$, entonces $\operatorname{erf} \left(\frac{\alpha}{\sigma\sqrt{2}} \right)$ es la probabilidad de que el error de una medición individual se encuentre en el intervalo $(-\alpha, \alpha)$ para un α positivo.

y para un nivel de significancia α se tiene que

$$\theta_W^* = F_{\theta_W}^{-1}(1 - \alpha), \quad (3.1.8)$$

donde $F_{\theta_W}^{-1}$ denota la función inversa de F_{θ_W} .

Cuando los campos generados bajo la hipótesis nula están correlacionados, normalmente no es posible encontrar una expresión aproximada para θ_W^* . Sin embargo, si uno tiene dos elementos estructurales V, W con $V \subset W$, en general se tiene que $\theta_W^* \leq \theta_V^*$, porque para cada voxel $\mathbf{x} \in L$, si $V_{\mathbf{x}} \subset W_{\mathbf{x}}$

$$\min_{\mathbf{y} \in W_{\mathbf{x}}} t(\mathbf{y}) \leq \min_{\mathbf{y} \in V_{\mathbf{x}}} t(\mathbf{y}), \quad (3.1.9)$$

y por consiguiente, $\theta_W \leq \theta_V$ para cada campo muestral t . En particular, si V es un voxel individual, $\theta_V^* = \theta_0^*$ (obtenemos el mismo umbral que con FWER). Esto significa que con SE grandes obtendremos umbrales más pequeños para un nivel de significancia dado, y así debería ser posible detectar RA con niveles de activación moderada, que no se detectan con el método FWER, pero cuya extensión es grande como para ocurrir por azar.

Potencia, Control del Error y Estimación de la Región de Activación

Considere nuevamente el caso donde los campos t están correlacionados espacialmente bajo H_0 , con $E[t(\mathbf{x})|H_0] = 0$ y $Var[t(\mathbf{x})|H_0] = 1$, y suponga que para voxeles en la región de activación A , $t(\mathbf{x})$ tiene la misma distribución que bajo la hipótesis nula H_0 , excepto por la adición de una constante positiva a , de manera que la relación señal/ruido (SNR, signal to noise ratio) sea igual a a . La potencia Q_1 de la prueba para detectar como activado un voxel \mathbf{x} , dado que $W_{\mathbf{x}}$ pertenece a la región de activación, se define como la probabilidad de que $\theta_W(\mathbf{x})$ (definido en ecuación (3.1.3)) sea mayor que θ_W^* , dado que $W_{\mathbf{x}} \subseteq A$, es decir

$$Q_1(W, a) = P(\theta_W(\mathbf{x}) > \theta_W^* | W_{\mathbf{x}} \subseteq A) \quad (3.1.10)$$

$$= P(t(\mathbf{y}) > \theta_W^*, \forall \mathbf{y} \in W_{\mathbf{x}} | W_{\mathbf{x}} \subseteq A) = q_1^{|W|}(W, a), \quad (3.1.11)$$

donde

$$q_1(W, a) = P(t(\mathbf{x}) > \theta_W^* | \mathbf{x} \in A) = P_{H_0}(t(\mathbf{x}) > \theta_W^* - a) = \Phi(\theta_W^* - a), \quad (3.1.12)$$

y Φ esta dada por la expresión (3.1.5). Observar que cuando $a \rightarrow \infty$, $Q_1 \rightarrow 1$. El poder de la prueba se incrementa con el tamaño del elemento estructural (dado que SE está contenido en RA). Esto muestra que para los voxeles $\mathbf{x} \in \mathbf{E}_W[A]$ con $|W| > 1$ la prueba con el SE W es notablemente más potente que las pruebas FWER, ya que corresponde a $|W| = 1$, siempre que la RA sea suficientemente grande como para que $\mathbf{E}_W[A]$ no sea vacía.

La probabilidad de detecciones falsas es:

$$q_0(W) = P_{H_0}(\theta_W(\mathbf{x}) > \theta_W^*) = P_{H_0}(t(\mathbf{y}) > \theta_W^*, \forall \mathbf{y} \in W_{\mathbf{x}}) = \Phi^{|W|}(\theta_W^*). \quad (3.1.13)$$

Notar que independientemente de la correlación espacial de t bajo H_0 , siempre se cumple que

$$P_{H_0}(\theta_W(\mathbf{x}) > \theta_W^*) \leq P_{H_0}(\theta_W > \theta_W^*) = 1 - F_{\theta_W}(\theta_W^*) = \alpha. \quad (3.1.14)$$

Para calcular la probabilidad de detección PD, esto es, la probabilidad de que $\theta_W(\mathbf{x}) > \theta_W^*$, cuando $W_{\mathbf{x}}$ está parcialmente contenida en A , se define la fracción de superposición β como:

$$\beta(\mathbf{x}) = \frac{|W_{\mathbf{x}} \cap A|}{|W|}. \quad (3.1.15)$$

Entonces para campos de ruido no correlacionado, se tiene:

$$PD(W, a, b) = P(\theta_W(\mathbf{x}) > \theta_W^* | \beta(\mathbf{x}) = b) = \left(q_0^{1-b}(W) q_1^b(W, a) \right). \quad (3.1.16)$$

La probabilidad de detección cae rápidamente cuando la fracción de superposición b decrece. Cuando la relación señal ruido a se incrementa, PD se incrementa también, pero la tasa de decaimiento con respecto a b sigue siendo aproximadamente la misma. Esto significa que PD será prácticamente igual a la probabilidad de detecciones falsas α , a menos que la proporción de voxeles activos b esté cercana a uno. Por esta razón, en la práctica se encuentra que la tasa de error empírica de falsos positivos de los procedimientos basados en morfología es similar a la obtenida con FWER, para la clase de SE que se usa en [16, Marroquín et al., (2011)]. Por otra parte, uno encuentra que el conjunto donde $\bar{E}_W[u_{\theta_W^*}[t]](\mathbf{x}) = 1$ está más cercano a la erosión morfológica de la RA A que a A misma, así que una aproximación más precisa a la forma de la RA real se puede obtener aplicando una dilatación morfológica a ese conjunto. Sin embargo, para una fracción de superposición dada, por ejemplo $b = 0.9$, el valor de PD crece con $|W|$. Esto significa que cuando $|W|$ es grande, la dilatación morfológica \mathbf{D}_W de $\mathbf{E}_W[u_{\theta_W^*}[t]]$ puede sobrestimar la extensión de A . Por esta razón, tiene sentido usar en general un SE $\tilde{W} \subset W$ más pequeño para llevar a cabo la dilatación, con el fin de conseguir una estimación más acertada de la región de activación. En otras palabras, el campo binario es igual a 1 solamente para aquellos voxeles que pertenecen a la RA en L y serán aproximados por el campo \hat{A}_W , el cual se define por:

$$\hat{A}_W = \mathbf{D}_{\tilde{W}} \left[\mathbf{E}_W \left[u_{\theta_W^*}[t] \right] \right]. \quad (3.1.17)$$

A $\mathbf{E}_W \left[u_{\theta_W^*}[t] \right]$ se le llama resultado de la detección de MBHT, y a \tilde{A}_W el estimador MBHT de A . Notar que si $\tilde{W} = W$, el operador $\mathbf{D}_{\tilde{W}} \mathbf{E}_W$ es igual al operador morfológico de apertura \mathbf{O}_W ([31, Serra, 1984]).

Cabe señalar que siempre y cuando la distribución conjunta de los valores de $t(\mathbf{x})$ fuera de RA sea la misma bajo H_0 y bajo $H_1 = (H_0)^c$, se tiene que

$$P \left(\hat{A}_W \subset \mathbf{D}_{\tilde{W}} \mathbf{D}_W[A] \right) \geq 1 - \alpha, \quad (3.1.18)$$

independientemente de la correlación espacial del ruido. Para ver esto es necesario notar que:

$$\begin{aligned}
 1 - P\left(\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] \subset \mathbf{D}_W[A]\right) &= P\left(\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] \not\subset \mathbf{D}_W[A]\right) \\
 &= P\left(\exists \mathbf{x} \notin \mathbf{D}_W[A]: \theta_W(\mathbf{x}) > \theta_W^*\right) \\
 &= P_{H_0}\left(\exists \mathbf{x} \notin \mathbf{D}_W[A]: \theta_W(\mathbf{x}) > \theta_W^*\right) \\
 &\leq P_{H_0}\left(\exists \mathbf{x} \in L: \theta_W(\mathbf{x}) > \theta_W^*\right) \leq \alpha,
 \end{aligned} \tag{3.1.19}$$

así que

$$\begin{aligned}
 1 - \alpha &\leq P\left(\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] \subset \mathbf{D}_W[A]\right) \\
 &\leq P\left(\mathbf{D}_{\tilde{W}} \left[\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] \right] \subset \mathbf{D}_{\tilde{W}}[\mathbf{D}_W[A]]\right),
 \end{aligned} \tag{3.1.20}$$

y de aquí se sigue la expresión (3.1.18). También hay que notar que siempre se tiene

$$\begin{aligned}
 P_{H_0}\left(\hat{A}_W \neq \emptyset\right) &= P_{H_0}\left(\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] \neq \emptyset\right) \\
 &= P_{H_0}\left(\exists \mathbf{x} \in L: \theta_W(\mathbf{x}) > \theta_W^*\right) \\
 &= P_{H_0}\left(\max_{\mathbf{x} \in L} \theta_W(\mathbf{x}) > \theta_W^*\right) = \alpha.
 \end{aligned} \tag{3.1.21}$$

Para describir en términos más precisos el tipo de control de error que se logra con MBHT, en [16, Marroquín, (2011)] se introducen las siguientes definiciones para un nivel de significancia α :

1. Un método tiene *Control de Error Basado en Cluster* (CBEC, Cluster-Based Error Control) si la probabilidad de declarar como activo un cluster conectado C tal que $C \cap A = \emptyset$ es menor que α .
2. Un método tiene *Control de Error de Localización* (LEC, Localization Error Control) acotado por ϵ , si la probabilidad de declarar activo algún voxel \mathbf{x} tal que $d(\mathbf{x}, A) > \epsilon$ es menor que α , donde $d(\mathbf{x}, A) = \min_{\mathbf{y} \in A} \|\mathbf{x} - \mathbf{y}\|$.

Es importante percatarse de que LEC acotado por ϵ implica control de FWER en el sentido fuerte sobre el conjunto definido por el complemento de la dilatación de la RA A por una esfera de radio ϵ .

La estimación del procedimiento MBHT tiene LEC acotado por la suma del radio de W más el radio de \tilde{W} . Si $\tilde{W} = W$, el procedimiento también tiene CBEC, ya que $\mathbf{E}_W \left[u_{\theta_W^*} [t] \right] (\mathbf{x}) = 1$ implica que con probabilidad $1 - \alpha$ el conjunto $W_{\mathbf{x}}$ tiene una intersección no vacía con la RA, y si $\tilde{W} = W$ en la ecuación (3.1.17) entonces el conjunto conectado $W_{\mathbf{x}}$ esta contenido en \hat{A}_W en su totalidad.

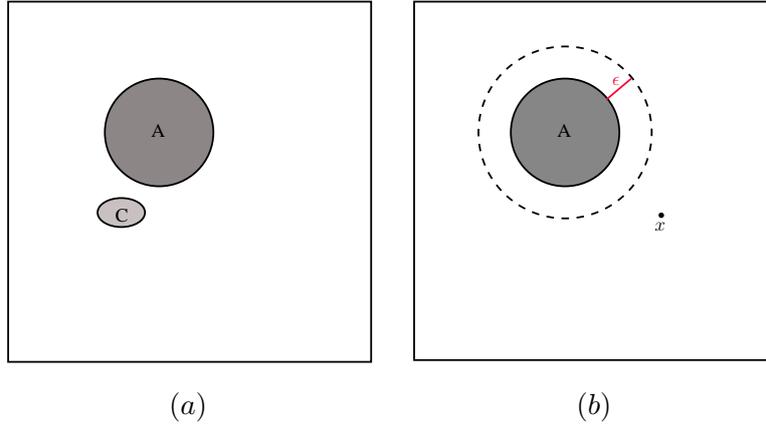


Fig. 3.1.1: (a) $P(C \text{ activo} | C \cap A = \emptyset) < \alpha$; (b) $P(x \text{ activo} | d(x, A) > \epsilon) < \alpha$.

3.1.2. Familias de Elementos Estructurales

Hasta ahora no hemos mencionado como realizar la selección de un SE apropiado cuando se está en una situación particular. Es habitual que no conozcamos a priori el tamaño ni la forma de las regiones de activación. Una solución razonable que proponen en [16, Marroquín, (2011)] es usar una familia de elementos estructurales $\mathcal{W} = \{W_1, \dots, W_N\}$ ² para calcular por separado los estimadores, mismos que se pueden combinar después. Sin embargo, aunque se pueden obtener umbrales individuales $\{\theta_{W_1}^*, \dots, \theta_{W_N}^*\}$ con el algoritmo 3.1.1, si simplemente se calcula el estimador para la región de activación como $\hat{A} = \bigcup_{j=1}^N \hat{A}_{W_j}$ dejaría de ser cierto que $P_{H_0}(\hat{A} \neq \emptyset) \leq \alpha$. En su lugar, en [16, Marroquín, (2011)] adoptan el enfoque descrito en [23, Pesarin, (2001)] (revisar en particular los capítulos 6 y 7) para combinar pruebas estadísticas, el cual consiste en transformar cada estadístico a la escala de significancia, y luego combinarla usando el máximo. Más específicamente, los estadísticos que se combinan son θ_{W_j} , $j = 1, \dots, N$, definidos por:

$$\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x}), \quad (3.1.22)$$

con $\theta_{W_j}(\mathbf{x})$ dado por la ecuación (3.1.3). Para transformar estos estadísticos a la escala de significancia, se usa la distribución $F_{\theta_{W_j}}$ para cada θ_{W_j} , bajo H_0 , y se define:

$$g(\mathbf{x}) = \max \left\{ F_{\theta_{W_1}}(\theta_{W_1}(\mathbf{x})), \dots, F_{\theta_{W_N}}(\theta_{W_N}(\mathbf{x})) \right\}, \quad (3.1.23)$$

$$g = \max_{\mathbf{x} \in L} g(\mathbf{x}). \quad (3.1.24)$$

²Por ejemplo, un conjunto de círculos o esferas concéntricas con radios incrementales donde W_1 es un voxel individual.

Sea g^* el umbral tal que $P_{H_0}(g > g^*) = \alpha$. Entonces, H_0 se rechaza si $g > g^*$ y los voxels responsables por rechazar H_0 son aquellos en los que $g(x) > g^*$. Observar que:

$$\begin{aligned} \{x \in L: g(x) > g^*\} &= \bigcup_{j=1}^N \{x \in L: F_{\theta_{W_j}}(\theta_{W_j}(x)) > g^*\} \\ &= \bigcup_{j=1}^N \{x \in L: \theta_{W_j}(x) > F_{\theta_{W_j}}^{-1}(g^*)\} \\ &= \bigcup_{j=1}^N \mathbf{E}_{W_j} \left[u_{\eta_{W_j}^*}[t] \right], \end{aligned} \quad (3.1.25)$$

donde el umbral $\eta_{W_j}^*$ está definido por:

$$\eta_{W_j}^* = F_{\theta_{W_j}}^{-1}(g^*). \quad (3.1.26)$$

Esto significa que se puede obtener un estimador de la región de activación controlando la probabilidad de error de falsos positivos con la siguiente expresión:

$$\hat{A} = \bigcup_{j=1}^N \tilde{A}_{W_j}, \quad (3.1.27)$$

donde

$$\tilde{A}_{W_j} = \mathbf{D}_{\tilde{W}_j} \left[\mathbf{E}_{W_j} \left[u_{\eta_{W_j}^*}[t] \right] \right], \quad (3.1.28)$$

y \tilde{W}_j, W_j se eligen de manera que disminuya la posibilidad de sobreestimar A , como se explicó anteriormente. En particular, si se tiene que $W_1 \subset W_2 \subset \dots \subset W_N$, se puede elegir algún índice $j_{\text{máx}} \leq N$, y hacer \tilde{W}_j igual a W_j si $j \leq j_{\text{máx}}$, e igual a $W_{j_{\text{máx}}}$ en otro caso. Notar que solamente haciendo $j_{\text{máx}} = N$ se puede garantizar que el procedimiento adaptable tendrá CBEC, pero con un $j_{\text{máx}}$ más pequeño se obtiene una cota más pequeña para los errores de localización, y en general una tasa de errores falsos positivos más pequeña. Desde luego que la mejor elección depende de la aplicación particular que se tenga. Para los experimentos realizados en [16, Marroquín, (2011)] se usa $j_{\text{máx}} = 2$.

La tasa de errores de falsos positivos es:

$$\begin{aligned} P_{H_0}(\hat{A} \neq \emptyset) &= P_{H_0} \left(\bigcup_{j=1}^N \mathbf{D}_{W_j} \mathbf{E}_{W_j} \left[u_{\eta_{W_j}^*}[t] \right] \neq \emptyset \right) \\ &= P_{H_0} \left(\bigcup_{j=1}^N \mathbf{E}_{W_j} \left[u_{\eta_{W_j}^*}[t] \right] \neq \emptyset \right) \\ &= P_{H_0}(g > g^*) = \alpha, \end{aligned} \quad (3.1.29)$$

y, al igual que antes, se tiene que $P(\hat{A}_W \subset \mathbf{D}_{\tilde{W}} \mathbf{D}_W[A]) \geq 1 - \alpha$.

3.2. Procedimiento Adaptable MBHT

Este procedimiento se basa en morfología matemática e incluye familias de elementos estructurales para obtener una estimación de la región de activación. Es adaptable en el siguiente sentido: en cada voxel $\mathbf{x} \in A$ el SE W_j apropiado realiza la detección, es decir, aquel para el cuál el estadístico $\theta_{W_j}(\mathbf{x})$ esta sobre el correspondiente umbral $\eta_{W_j}^*$. El método se divide en dos etapas. Primero, se calculan los umbrales $\eta_{W_j}^*$ para todo $W_j \in \mathcal{W}$. Esta parte sólo se ejecuta una sola vez para cada clase de campos de prueba que comparten el modelo de ruido M_0 . El algoritmo 3.2.1 muestra el procedimiento de esta etapa.

Algorithm 3.2.1: MBHT Adaptable (Primer etapa)

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α y una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de SEs.

Output: Un conjunto de umbrales de detección $\{\eta_{W_j}^*\}$.

1 Generar muestras de campos t usando M_0 .

2 **foreach** *campo muestral* t **do**

3 **foreach** $W_j \in \mathcal{W}$ **do**

4 Calcular

$$\theta_{W_j}(\mathbf{x}) = \min_{\mathbf{y} \in W_{j\mathbf{x}}} t(\mathbf{y}) \quad (3.2.1)$$

 donde $W_{j\mathbf{x}}$ denota la translación del conjunto W_j a la posición \mathbf{x} .

5 Calcular la variable aleatoria

$$\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x}). \quad (3.2.2)$$

6 **foreach** $W_j \in \mathcal{W}$ **do**

7 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} .

8 **foreach** *campo muestral* t **do**

9 **foreach** $W_j \in \mathcal{W}$ y cada $\mathbf{x} \in L$ **do**

10 calcular las variables aleatorias: $\theta_{W_j}(\mathbf{x})$ con (3.2.1) y $g(\mathbf{x})$ con (3.1.23)

11 Calcular g con la ecuación (3.1.24).

12 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .

13 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.

14 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales $\eta_{W_j}^*$ usando la ecuación (3.1.26).

15 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.

Como puede observarse, en la línea 1 se obtiene un conjunto de campos muestrales que satisfacen H_0 ; en las líneas 2 a 7 se obtiene la correspondiente distribución empírica de los estadísticos

θ_{W_j} , en las líneas 8 a 12 se calcula la distribución empírica del máximo de estos estadísticos en la escala de significancia (variable g) y finalmente en la línea 15 se obtienen los correspondientes umbrales finales.

En la segunda etapa de Adaptive MBHT, se estima la región de activación RA para una imagen de prueba dada. A continuación se muestra su algoritmo.

Algorithm 3.2.2: Adaptive MBHT (Segunda etapa)

Input: Un campo de prueba t , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de SEs, y un conjunto correspondiente de umbrales de detección $\{\eta_{W_j}^*\}$.

Output: Una estimación de RA \hat{A} en t .

- 1 Para cada $W_j \in \mathcal{W}$ calcular los campos \tilde{A}_{W_j} con la ecuación (3.1.28).
 - 2 Calcular el estimador \hat{A} usando la ecuación (3.1.27).
 - 3 **return** \hat{A} .
-

Al igual que con un elemento estructural individual, es posible trazar la detección de p -valores en cada localización para la familia \mathcal{W} . En este caso, el p -valor localizado en \mathbf{x} está dado por:

$$p(\mathbf{x}) = 1 - F_g(g(\mathbf{x})). \quad (3.2.3)$$

Notar que con los umbrales calculados $\eta_{W_j}^*$ con la ecuación (3.1.26), cualquier familia W dará resultados con un control del error apropiado, aunque la habilidad de cada familia W para recuperar formas específicas variara. En general, es una buena idea incluir en \mathcal{W} SEs relativamente grandes, de esta forma se pueden detectar regiones grandes con activaciones moderadas; también es buena idea incluir SEs pequeños, de esta manera es posible recuperar formas más precisas siempre que sus niveles de activación sean lo suficientemente altos. En particular, W_1 debería ser igual a la RA más pequeña posible, la cuál usualmente es de un voxel. Se debería tener en cuenta que el proceso completo de estimación tiene LEC acotado por el radio de W_N más el radio de \tilde{W}_N (el cual es igual a 1 voxel si $j_{\text{máx}} = 2$). En [16, Marroquín, (2011)] utilizan W_N con un radio igual a 5 voxeles, pero en general la elección depende de la aplicación que se tenga en particular.

3.3. Multi-Threshold Supra Threshold Cluster Size (MTSTCS)

Debido a que el desempeño de los métodos basados en técnicas STCC depende fuertemente de la selección del umbral apropiado, en el artículo [16, Marroquín, (2011)] utilizan las ideas clave del procedimiento Adaptive MBHT para definir un procedimiento adaptable que se basa en el algoritmo STCS (ver sección 2.2). Esto se hace recurriendo a múltiples umbrales de intensidad y combinando los resultados de las pruebas parciales usando la función máxima en la escala de

significancia (ver sección 3.1.2). A este método le llamaron *Multi-Threshold STCS* (MTSTCS). Para implementar este enfoque basta con realizar algunas modificaciones al algoritmo 3.2.1 para obtener el siguiente:

Algorithm 3.3.1: MTSTCS (Primer etapa)

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α y una secuencia incremental de umbrales de intensidad

$$\mathcal{W} = \{W_1, \dots, W_N\}.$$

Output: Un conjunto de umbrales de área $\{\eta_{W_j}^*\}$.

- 1 Generar muestras de campos t usando M_0 .
 - 2 **foreach** *campo muestral* t **do**
 - 3 **foreach** $W_j \in \mathcal{W}$ **do**
 - 4 Calcular el estadístico $\theta_{W_j}(\mathbf{x})$ como el área del cluster conectado en el campo $u_{W_j}[t]$ al que pertenece \mathbf{x} .
 - 5 Calcular la variable aleatoria $\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x})$.
 - 6 **foreach** $W_j \in \mathcal{W}$ **do**
 - 7 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} .
 - 8 **foreach** *campo muestral* t **do**
 - 9 **foreach** $W_j \in \mathcal{W}$ *y cada* $\mathbf{x} \in L$ **do**
 - 10 Calcular $\theta_{W_j}(\mathbf{x})$ como en la línea 4.
 - 11 Calcular $g(\mathbf{x})$ con la ecuación (3.1.23).
 - 12 Calcular g con la ecuación (3.1.24).
 - 13 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .
 - 14 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.
 - 15 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales de área $\eta_{W_j}^*$ usando la ecuación (3.1.26).
 - 16 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.
-

Para la segunda etapa del algoritmo también es necesario realizar algunos cambios.

Algorithm 3.3.2: MTSTCS (Segunda etapa)

Input: Un campo de prueba t , una secuencia incremental de umbrales de intensidad

$\mathcal{W} = \{W_1, \dots, W_N\}$, y un conjunto correspondiente de umbrales de área $\{\eta_{W_j}^*\}$.

Output: Una estimación de RA \hat{A} en t .

- 1 Para cada $W_j \in \mathcal{W}$ calcular los campos \tilde{A}_{W_j} como la colección de todos los clusters conectados en $u_{W_j}[t]$ cuya área es mayor que $\eta_{W_j}^*$.
 - 2 Calcular el estimador \hat{A} como la unión de las estimaciones \tilde{A}_{W_j} (usando la ecuación (3.1.27)).
 - 3 **return** \hat{A} .
-

Aunque MTSTCS supera en parte algunas de las limitaciones del método estándar STCS, como lo es el problema de la selección de un umbral adecuado, sigue teniendo el problema que se deriva del hecho de que la conectividad de voxeles no refleja bien el concepto intuitivo de “un clúster compacto contiguo”. Este método tiene control débil del error de falsos positivos y CBEC pero no tiene LEC (ver sección 3.1.1).

3.4. Multi-Threshold Supra Threshold Cluster Mass (MTSTCM)

Al igual que en la sección anterior, es posible aplicar el mismo procedimiento que se usó para desarrollar MTSTCS, para generalizar el enfoque propuesto por [4, Bullmore et al., (1999)]. De esta manera surge el método *Multi-Threshold Supra Threshold Cluster Mass* (MTSTCM).

Este método también utiliza una sucesión creciente de umbrales de intensidad para minimizar el problema de la elección de un umbral apropiado que se tiene en el procedimiento estándar de STCM. El único cambio que se realiza al método MTSTCS consiste en cambiar el estadístico de prueba. Es decir, que en lugar de calcular el área del clúster C al que pertenece \mathbf{x} , se calcula la masa del clúster C al que pertenece \mathbf{x} como se muestra en el algoritmo 3.4.1.

Algorithm 3.4.1: MTSTCM (Primer etapa)

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α y una secuencia incremental de umbrales de intensidad $\mathcal{W} = \{W_1, \dots, W_N\}$.

Output: Un conjunto de umbrales de masa $\{\eta_{W_j}^*\}$.

- 1 Generar muestras de campos t usando M_0 .
- 2 **foreach** *campo muestral* t **do**
- 3 **foreach** $W_j \in \mathcal{W}$ **do**
- 4 Calcular el estadístico $\theta_{W_j}(\mathbf{x})$ como la masa del cluster C conectado en el campo $u_{W_j}[t]$ al que pertenece \mathbf{x} con la ecuación:

$$\theta_{W_j}(\mathbf{x}) = \sum_{\mathbf{x} \in C} h(\mathbf{x}) \quad \text{donde} \quad h(\mathbf{x}) = \begin{cases} t(\mathbf{x}) - W_j, & \text{si } t(\mathbf{x}) > W_j; \\ 0, & \text{si } t(\mathbf{x}) \leq W_j. \end{cases} \quad (3.4.1)$$
- 5 Calcular la variable aleatoria $\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x})$.
- 6 **foreach** $W_j \in \mathcal{W}$ **do**
- 7 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} .
- 8 **foreach** *campo muestral* t **do**
- 9 **foreach** $W_j \in \mathcal{W}$ y *cada* $\mathbf{x} \in L$ **do**
- 10 Calcular $\theta_{W_j}(\mathbf{x})$ como en la línea 4.
- 11 Calcular $g(\mathbf{x})$ con la ecuación (3.1.23).
- 12 Calcular g con la ecuación (3.1.24).
- 13 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .
- 14 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.
- 15 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales de área $\eta_{W_j}^*$ usando la ecuación (3.1.26).
- 16 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.

En cuanto a la segunda etapa sólo es necesario cambiar la línea 1 del algoritmo 3.3.2 por el siguiente enunciado: para $W_j \in \mathcal{W}$ calcular los campos \tilde{A}_{W_j} como la colección de todos los clusters conectados en $u_{W_j}[t]$ cuya *masa* sea mayor que $\eta_{W_j}^*$, donde la masa en un voxel \mathbf{x} se obtiene con la expresión de la línea 4 del algoritmo 3.4.1. El estimador final \hat{A} también se obtiene con la fórmula (3.1.27). Es importante hacer notar que este método, al igual que MTSTCS, tampoco cuenta con control del error de localización.

3.5. MBHT-TFCE

Con el método de *Threshold Free Cluster Enhancement* (TFCE), descrito en la sección 2.4, en general se pueden obtener resultados muy buenos y estables. Pero ¿por qué sucede esto? Con las técnicas mostradas en las secciones 3.3 y 3.4 es muy posible que se produzcan clusters conectados porosos como resultado al utilizar umbrales de intensidad muy bajos. Por el contrario, con TFCE, dichos clusters reciben un peso relativamente pequeño $(h_j - h_0)^2$ en la puntuación combinada.

Debido a la observación anterior, en [16, Marroquín et al, (2011)] se efectúa un ajuste del procedimiento TFCE usando MBHT, de modo que la extensión del cluster no se mide por el área de conjuntos de voxeles conectados, sino por un criterio morfológico que, como se puntualizó anteriormente, manifiesta la noción de un clúster compacto contiguo de una mejor manera.

Para describir el proceso MBHT-TFCE, suponga que se tiene una familia creciente de SEs W_1, \dots, W_M dada y un umbral de intensidad h_j . Suponga también que se puede definir $e(\mathbf{x}, h_j)$ como el tamaño del SE más grande de la familia, digamos W_j , tal que \mathbf{x} está contenido en W_j para algún \mathbf{y} en la vecindad de \mathbf{x} , y todos los voxeles en $W_{j\mathbf{y}}$ tienen el valor 1 en $u_{h_j}[t]$. Esto es equivalente a fijar $e(\mathbf{x}, h_j)$ como el tamaño del SE más grande tal que la apertura morfológica realizada con ese SE en el campo $u_{h_j}[t]$ tiene el valor 1 en el voxel \mathbf{x} . En términos más precisos

$$e(x, h) = \begin{cases} 0, & \text{si } u_{h_j}[t] = 0; \\ W_{kmax}, & \text{en otro caso;} \end{cases} \quad (3.5.1)$$

donde

$$kmax = \max_x \mathbf{O}_{W_k} [u_{h_k}[t]] (\mathbf{x}) = 1. \quad (3.5.2)$$

Se recomienda que los umbrales sean los valores $\eta_{W_j}^*$, para $j = 1, \dots, N$, y el umbral de referencia h_0 sea $h_0 = \eta_{W_N}^* - 0.1|\eta_{W_N}^*|$. Al igual que en [32, Smith and Nichols, (2009)], el valor crítico para el estadístico $TFCE(\mathbf{x})$ dado por la ecuación (2.4.1) se puede obtener de manera no paramétrica usando procedimientos de permutación.

Métodos Basados en Estimadores Robustos para Detección de Activaciones en Campos Aleatorios Discretos

Con el fin de diseñar nuevos algoritmos que permitan la detección de activaciones en campos aleatorios discretos, se buscaron operadores robustos, para calcular estadísticos de prueba que cumplieran con las siguientes propiedades:

1. Soporte local, para garantizar control del error de localización.
2. Distribución del estadístico de prueba bajo la hipótesis nula lo más concentrada posible, para que los umbrales de detección de activaciones sean menores.
3. Robustez con respecto a los efectos de bordes, para evitar errores (falsos positivos y falsos negativos) en las fronteras de la región de activación. Esto al menos en regiones de activación convexas.

En las figura 4.0.1 a 4.0.4 podemos ver que si elegimos como estadísticos de prueba a la moda, mediana o media en una vecindad del voxel x del campo t (el cual ha sido generado bajo la hipótesis nula), obtenemos un histograma de menor varianza y por ende de distribución más “angosta”. Por esta razón usamos los operadores con soporte local de media, mediana y moda. También consideramos el filtro bilateral ya que es un promedio ponderado (ver sección 4.2.3) que toma en cuenta la intensidad de los voxeles vecinos.

Al considerar una vecindad del voxel $x \in A$, con una proporción pequeña de sus vecinos fuera de la región de activación, éstos podrían considerarse como datos atípicos y viceversa. Es sabido que los estadísticos de orden son más robustos contra “outliers” en comparación con las medidas de tendencia central. Por ello, es interesante observar cual es el comportamiento de estos estadísticos en la frontera de la región de activación.

Utilizando las ideas anteriores, a continuación se presentan seis métodos nuevos para la detección de activaciones en campos aleatorios discretos. Para su desarrollo se emplean algunos

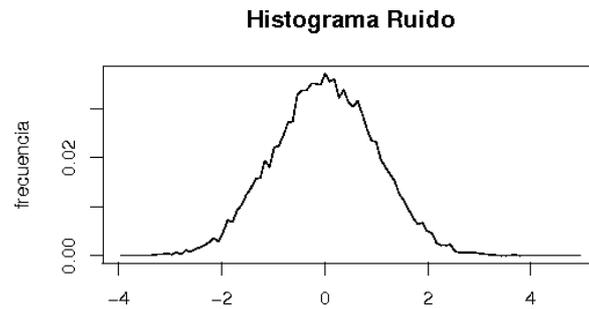


Fig. 4.0.1: Histograma de ruido. Esto es, histograma del campo t generado bajo la hipótesis nula.

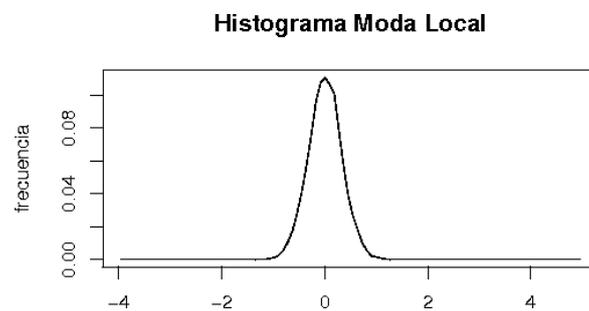


Fig. 4.0.2: Histograma de la Moda, con soporte local la circunferencia de radio $r = 4$, del campo t . Con t generado bajo la hipótesis nula.

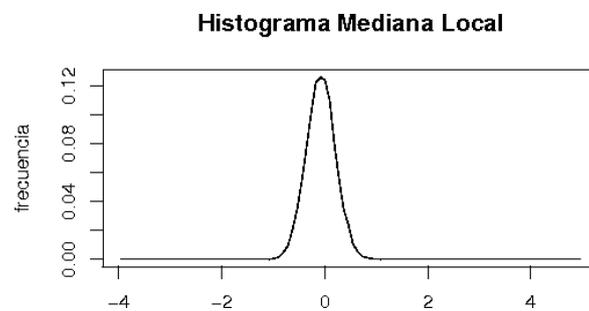


Fig. 4.0.3: Histograma de la Mediana, con soporte local la circunferencia de radio $r = 4$, del campo t . Con t generado bajo la hipótesis nula.

conceptos del procedimiento adaptable MBHT visto en la sección 3.2.

4.1. Método Basado en Estadísticos de Orden

Suponga que se tiene una muestra aleatoria dada X_1, X_2, \dots, X_m de una población, entonces podemos obtener el conjunto $(X_{1,m}, X_{2,m}, \dots, X_{m,m})$ formado por los mismos elementos de la muestra puestos en orden creciente, esto es, $X_{1,m} \leq X_{2,m} \leq \dots \leq X_{m,m}$. También es común encontrarnos con la notación $X_{(h)} = X_{h,m}$, que sólo indica el lugar dentro de la ordenación, pero

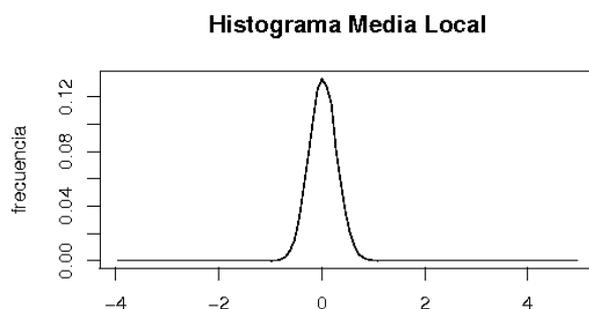


Fig. 4.0.4: Histograma de la Media, con soporte local la circunferencia de radio $r = 4$, del campo t . Con t generado bajo la hipótesis nula.

no el tamaño de la muestra.

Definición 4.1.1. *A la variable aleatoria $X_{(h)}$ se le llama estadístico de orden h de la muestra. En general a $X_{(1)}, \dots, X_{(m)}$ se les llama estadísticos de orden.*

Luego, el primer estadístico de orden $X_{(1)}$ y el último $X_{(m)}$, son el mínimo y el máximo de la muestra respectivamente. Además, la mediana muestral esta dada por:

$$\begin{cases} X_{(\frac{m+1}{2})}, & \text{si } m \text{ es impar,} \\ \frac{X_{(\frac{m}{2})} + X_{(\frac{m}{2}+1)}}{2}, & \text{si } m \text{ es par.} \end{cases} \quad (4.1.1)$$

En el algoritmo 4.1.1 se especifican los pasos del método basado en estadísticos de orden.

Algorithm 4.1.1: Estadísticos de Orden

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de SEs y un entero n con $1 \leq n \leq \frac{|W_N| + 1}{2}$.

Output: Un conjunto de umbrales $\{\eta_{W_j}^*\}$.

1 Generar muestras de campos t usando M_0 .

2 **foreach** *campo muestral* t **do**

3 **foreach** $W_j \in \mathcal{W}$ **do**

4 Calcular

$$\theta_{W_j}(\mathbf{x}) = t_{n, |W_j|}(\mathbf{x}) = \mathbf{n}^\circ \min_{\mathbf{y} \in W_{j\mathbf{x}}} t(\mathbf{y}) \quad (4.1.2)$$

donde $\mathbf{n}^\circ \min_{\mathbf{y} \in W_{j\mathbf{x}}} t(\mathbf{y})$ denota al estadístico de orden n del conjunto $W_{j\mathbf{x}}$, si $n < \frac{|W_j| + 1}{2}$ y es igual a la mediana de $W_{j\mathbf{x}}$ en otro caso.

5 Calcular la variable aleatoria

$$\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x}) \quad (4.1.3)$$

6 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} para cada $W_j \in \mathcal{W}$.

7 **foreach** *campo muestral* t **do**

8 **foreach** $W_j \in \mathcal{W}$ y cada $\mathbf{x} \in L$ **do**

9 Calcular $\theta_{W_j}(\mathbf{x})$ con (4.1.2).

10 Calcular $g(\mathbf{x})$ con la ecuación (3.1.23).

11 Calcular g con la ecuación (3.1.24).

12 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .

13 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.

14 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales $\eta_{W_j}^*$ usando la ecuación (3.1.26).

15 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.

Como puede apreciarse con la expresión 4.1.2 del algoritmo 4.1.1, la variable $\theta_{W_j}(\mathbf{x})$ se define como el estadístico de orden n del elemento estructural $W_{j\mathbf{x}}$ (SE W_j trasladado al voxel \mathbf{x}) si n es menor que $\frac{|W_j| + 1}{2}$ ó menor que $\frac{|W_j|}{2}$ (esto es, menor a la posición de la mediana dependiendo de si $|W_j|$ es impar o par respectivamente); en caso contrario hacemos $\theta_{W_j}(x)$ igual al valor de la mediana muestral con la expresión (4.1.1). El método MBHT adaptable puede verse como un caso particular del proceso basado en estadísticos de orden, ya que se trata del estadístico de orden $n = 1$.

Para la segunda parte del algoritmo es necesario redefinir el operador erosión \mathbf{E}_W como

$$\bar{\mathbf{E}}_W[B](\mathbf{x}) = \mathfrak{n}^\circ \min_{\mathbf{y} \in W_{\mathbf{x}}} B(\mathbf{y}) = B_{n, |W_{\mathbf{x}}|}(\mathbf{x}), \quad (4.1.4)$$

esto es, como el estadístico de orden n del SE $W_{\mathbf{x}}$ del campo binario $B(\mathbf{x})$. Luego el algoritmo de la segunda etapa queda como 4.1.2:

Algorithm 4.1.2: Estadísticos de Orden (Segunda etapa)

Input: Un campo de prueba t , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de SEs, y un conjunto correspondiente de umbrales de detección $\{\eta_{W_j}^*\}$.

Output: Una estimación de RA \hat{A} en t .

1 Para cada $W_j \in \mathcal{W}$ calcular los campos \tilde{A}_{W_j} como

$$\tilde{A}_{W_j} = \bar{\mathbf{E}}_{W_j} \left[u_{\eta_{W_j}^*} [t] \right]. \quad (4.1.5)$$

2 Calcular el estimador $\hat{A} = \bigcup_{j=1}^N \tilde{A}_{W_j}$.

3 **return** \hat{A} .

4.2. Métodos Basados en el Histograma Local Ponderado

Para los métodos siguientes es conveniente calcular el histograma local ponderado en el voxel \mathbf{x} asociado a un elemento estructural $W_{\mathbf{x}}$. Esto nos permitirá encontrar de forma sencilla la moda local y calcular el filtro bilateral; además podemos obtener la distribución local en $W_{\mathbf{x}}$ y con esta última podemos hallar la mediana local o el cuantil (local) que se desee. A continuación se explica la manera en que se calcula dicho histograma.

Suponga que se tiene un modelo estadístico M_0 con el cual generamos un campo aleatorio t bajo la hipótesis nula. Usaremos $\mathbf{x} = (x_1, x_2)$ para identificar un voxel, y además suponga que se tienen $m \times n$ voxeles. Lo primero que hacemos es discretizar el rango de valores que toma $t(\mathbf{x})$ (rango de intensidad de la imagen) en B intervalos (bins). Representaremos a cada elemento del intervalo i con el punto medio de éste, el cuál será denotado por b_i y, llamaremos a l_i, l_{i+1} los extremos inferior y superior del intervalo i respectivamente. Esto es

$$b_i = \frac{l_i + l_{i+1}}{2}.$$

Además definimos la función $\delta(\mathbf{x}, i)$ como:

$$\delta(\mathbf{x}, i) = \begin{cases} 1, & \text{si } t(\mathbf{x}) \in [l_i, l_{i+1}); \\ 0, & \text{en otro caso.} \end{cases} \quad (4.2.1)$$

Observar que con la definición (4.2.1), al evaluar δ para todos los voxeles y todos los intervalos, obtenemos, en el caso de una imagen, un “cubo” de elementos binarios, al que denotaremos

como el cubo δ . Por conveniencia, algunas veces denotaremos la función δ con $\delta_i(\mathbf{x})$ para hacer énfasis en que los voxeles son la variable de interés, o la denotaremos con $\delta_x(i)$ para señalar que la variable de interés son los intervalos; sin perder de vista que nos referimos a la misma función δ .

Usando la notación anterior podemos obtener el histograma local en \mathbf{x} asociado al SE W . Para ello definimos la frecuencia del intervalo i , $h_{i,W}(\mathbf{x})$ como: el número de veces en que la intensidad de un voxel que pertenece a W_x cae dentro de este intervalo, esto es

$$h_{i,W}(\mathbf{x}) = \sum_{\mathbf{y} \in W_x} \delta_i(\mathbf{y}). \quad (4.2.2)$$

Entonces definimos el histograma local en \mathbf{x} como

$$h_{\mathbf{x},W}(i) = h_{i,W}(\mathbf{x}), \quad (4.2.3)$$

para $i \in \{1, \dots, B\}$.

Con la finalidad de dar mayor importancia a la información que aportan los voxeles que se encuentran más cerca del voxel \mathbf{x} , asignaremos pesos dependiendo de su distancia a \mathbf{x} . Para esto usaremos una Gaussiana bidimensional centrada en \mathbf{x} y con matriz de covarianzas igual a un múltiplo de la identidad. Por lo tanto reescribiremos $h_{i,W}(\mathbf{x})$ como:

$$\hat{h}_{i,W}(\mathbf{x}) = \sum_{\mathbf{y} \in W_x} \delta_i(\mathbf{y}) w(\mathbf{x} - \mathbf{y}), \quad (4.2.4)$$

donde

$$w(\mathbf{x}) = ce^{-\frac{1}{2} \frac{\|\mathbf{x}\|^2}{\sigma^2}}, \quad (4.2.5)$$

con c constante de normalización tal que $\sum_{\mathbf{y} \in W_x} w(\mathbf{x} - \mathbf{y}) = 1$. Recordar que $\mathbf{y} \in W_x$ implica que $\|\mathbf{y} - \mathbf{x}\| = \|\mathbf{x} - \mathbf{y}\| \leq r$ para un radio r dado. Si se consideran los voxeles \mathbf{y} tales que $w(\mathbf{y} - \mathbf{x})$ sea igual a una constante, entonces $\|\mathbf{y} - \mathbf{x}\|$ también es igual a una constante, luego las curvas de nivel de $w(\mathbf{y} - \mathbf{x})$ son circunferencias con centro en \mathbf{x} .

La distribución acumulada local en \mathbf{x} la podemos calcular con la siguiente expresión:

$$\hat{H}_{\mathbf{x},W}(i) = \sum_{k=0}^i \hat{h}_{k,W}(\mathbf{x}). \quad (4.2.6)$$

Notar que, si en la ecuación (4.2.4) tomamos $\mathbf{z} = \mathbf{x} - \mathbf{y}$, entonces $\mathbf{y} = \mathbf{x} - \mathbf{z}$; luego al sustituir estas variables en la ecuación (4.2.4), obtenemos:

$$\hat{h}_{i,W}(\mathbf{x}) = \sum_{\mathbf{z} \in W} w(\mathbf{z}) \delta_i(\mathbf{x} - \mathbf{z}). \quad (4.2.7)$$

Entonces, para un voxel particular \mathbf{x}_0 , si variamos i en (4.2.7) obtenemos el histograma local ponderado en \mathbf{x}_0 asociado a W , $\hat{h}_{\mathbf{x}_0,W}(i)$. En cambio, si en lugar de fijar un voxel fijamos un

intervalo i_0 de intensidad b_{i_0} , mientras variamos el voxel \mathbf{x} , la expresión anterior representa la convolución entre el campo binario de nivel i_0 (imagen binaria de dimensión $m \times n$ correspondiente al intervalo i_0) y el kernel Gaussiano w . En la práctica, cuando el voxel \mathbf{x} se encuentra en los bordes de la imagen, algunos de los voxeles $\mathbf{y} \in W_{\mathbf{x}}$ quedan fuera de la imagen, para eludir el problema realizaremos la convolución circular¹, esto es, $(\delta_{i_0} \otimes w)(\mathbf{x})$. Observar que se tienen B niveles (intervalos) y para cada uno de ellos se realiza la convolución (4.2.7). De lo anterior notamos que calcular el histograma local ponderado en cada voxel $\mathbf{x} \in L$ asociado al elemento estructural W de radio r , es equivalente a suavizar cada campo binario de nivel i (de los B intervalos que se tienen) usando un kernel Gaussiano $w(\mathbf{x})$ tal que $\sum_{\|\mathbf{x}\| \leq r} w(\mathbf{x}) = 1$. La equivalencia anterior hace más eficiente al método ya que el operador Gaussiano $w(\mathbf{x})$ es separable. Para ver esto notar que

$$\begin{aligned} \hat{h}_{i,W}(\mathbf{x}) &= (\delta_i \otimes w)(\mathbf{x}) \\ &= \sum_{\mathbf{y} \in W_{\mathbf{x}}} \delta_i(\mathbf{y})w(\mathbf{x} - \mathbf{y}), \end{aligned}$$

como cada voxel $\mathbf{x} = (x_1, x_2)$, la expresión anterior se convierte en

$$\begin{aligned} \hat{h}_{i,W}(x_1, x_2) &= (\delta_i \otimes w)(x_1, x_2) \\ &= \sum_{y_1} \sum_{y_2} \delta_i(y_1, y_2)w(x_1 - y_1, x_2 - y_2), \end{aligned}$$

donde $\|(x_1 - y_1, x_2 - y_2)\| \leq r$ y w como en (4.2.5). Entonces

$$\begin{aligned} \hat{h}_{i,W}(x_1, x_2) &= \sum_{y_1} \sum_{y_2} \delta_i(y_1, y_2)ce^{-\frac{1}{2}\frac{(x_1-y_1)^2+(x_2-y_2)^2}{\sigma^2}} \\ &= \sum_{y_1} \sum_{y_2} \delta_i(y_1, y_2)ce^{-\frac{1}{2}\left(\frac{(x_1-y_1)^2}{\sigma^2} + \frac{(x_2-y_2)^2}{\sigma^2}\right)} \\ &= \sum_{y_1} \sum_{y_2} \delta_i(y_1, y_2)ce^{\left(-\frac{1}{2}\frac{(x_1-y_1)^2}{\sigma^2} - \frac{1}{2}\frac{(x_2-y_2)^2}{\sigma^2}\right)} \\ &= \sum_{y_1} \sum_{y_2} \delta_i(y_1, y_2)\sqrt{ce^{-\frac{1}{2}\frac{(x_1-y_1)^2}{\sigma^2}}}\sqrt{ce^{-\frac{1}{2}\frac{(x_2-y_2)^2}{\sigma^2}}} \\ &= \sum_{y_1} \left(\sum_{y_2} \delta_i(y_1, y_2)w_{x_2}(x_2 - y_2) \right) w_{x_1}(x_1 - y_1) \\ &= \sum_{y_1} (\delta_i \otimes w_{x_2})(y_1, x_2) w_{x_1}(x_1 - y_1) \\ &= (\delta_i \otimes w_{x_2} \otimes w_{x_1})(x_1, x_2), \end{aligned}$$

con $w_{x_1}(x) = w_{x_2}(x) = \sqrt{ce^{-\frac{1}{2}\frac{x^2}{\sigma^2}}}$. Esto muestra que convolucionar la imagen binaria δ_i con el

¹Al utilizar la convolución circular, extendemos la imagen δ_{i_0} periódicamente.

kernel w es equivalente a convolucionar la imagen primero por renglones con el kernel w_{x_2} y después por columnas con el kernel w_{x_1} .

A partir de aquí cambiaremos la familia de elementos estructurales $\mathcal{W} = \{W_1, W_2, \dots, W_N\}$ que hemos estado usando por una familia de Gaussianas. Así, en lugar de que W_j sea un círculo de radio r_j como antes, ahora será el conjunto de elementos que pertenecen al kernel Gaussiano de varianza σ_j^2 , con la secuencia incremental $\sigma_1, \sigma_2, \dots, \sigma_N$ de varianzas, y w_j será el conjunto de valores del kernel de cada uno de los puntos que pertenecen a éste. En la práctica, se toma un kernel Gaussiano truncado, de manera que el kernel de varianza σ_j^2 esté totalmente contenido en el círculo de radio $r_j = 4 * \sigma_j$.

Una vez que hemos calculado el histograma local ponderado en \mathbf{x} , $\hat{h}_{\mathbf{x},W}(i)$, le aplicamos un suavizador Gaussiano de varianza σ_I^2 adecuada² para atenuar los efectos causados por el ruido. El resultado de este procedimiento es el histograma local ponderado suavizado $\tilde{h}_W(\mathbf{x}, i)$. La forma de obtenerlo es calculando la convolución entre $\hat{h}_W(\mathbf{x}, i)$ y el operador Gaussiano dado por

$$\mathcal{K}_{\mathbf{x}}(i) = \mathcal{K}(\mathbf{x}, i) = d \exp\left(-\frac{1}{2} \frac{i^2}{\sigma_I^2}\right), \quad (4.2.8)$$

donde d es una constante de normalización. Entonces podemos expresar la convolución como

$$\left(\hat{h}_{\mathbf{x},W} * \mathcal{K}_{\mathbf{x}}\right)(i) = \sum_{k \in \Omega} \hat{h}_{\mathbf{x},W}(i - k) \mathcal{K}_{\mathbf{x}}(k), \quad (4.2.9)$$

donde Ω es el soporte del operador Gaussiano \mathcal{K} .

Lo importante es observar que estamos aplicando un operador Gaussiano por capa, el cuál es separable, y luego un operador Gaussiano para el histograma local ponderado. Entonces podemos calcular $\tilde{h}_W(\mathbf{x}, i)$ para cada voxel $\mathbf{x} \in L$ asociado a un elemento W con

$$\begin{aligned} \tilde{h}_W(\mathbf{x}, i) &= \tilde{h}_W(x_1, x_2, i) \\ &= \left(\hat{h}_{x,W} * \mathcal{K}_x\right)(i) \\ &= \left(\hat{h}_W * \mathcal{K}\right)(x_1, x_2, i) \\ &= (\delta \otimes w_{x_2} \otimes w_{x_1} * \mathcal{K})(x_1, x_2, i) \\ &= (\delta * \mathcal{K} \otimes w_{x_2} \otimes w_{x_1})(x_1, x_2, i). \end{aligned} \quad (4.2.10)$$

De lo anterior podemos decir que estamos convolucionando con un operador Gaussiano cada dirección del cubo δ , por renglón, columna y altura. Además, si $t(\mathbf{x}_0) \in [l_j, l_{j+1})$, entonces:

$$\delta_{\mathbf{x}_0}(i) = \begin{cases} 1, & \text{si } i = j; \\ 0, & \forall i \in \{1, \dots, B\} \text{ con } i \neq j; \end{cases} \quad (4.2.11)$$

²Desde luego que surge la pregunta de como elegir “adecuadamente” el valor de esa varianza, pero por ahora lo dejaremos como un parámetro que será necesario entonar. Ver capítulo de Experimentos y Resultados.

luego $\delta_{\mathbf{x}_0}(j) = 1$ indica un impulso en j . Además se debe tener presente que la convolución es conmutativa, por lo que podríamos realizar primero la convolución de δ con \mathcal{K} y después por capa (separando en convolución por renglón w_{x_2} y por columna w_{x_1}).

El resultado de la convolución

$$\begin{aligned} (\delta * \mathcal{K})(x_1, x_2, i) &= \sum_{k \in \Omega} \mathcal{K}(x_1, x_2, k) \delta(x_1, x_2, i - k) \\ &= \sum_{l \in \Omega_i} \mathcal{K}(x_1, x_2, i - l) \delta(x_1, x_2, l) && \text{con } l = i - k \\ &= \sum_{l \in \Omega_i} \delta(x_1, x_2, l) \mathcal{K}(x_1, x_2, i - l), \end{aligned} \quad (4.2.12)$$

donde $\Omega_i = \{i - k | k \in \Omega\}$, es el kernel \mathcal{K} trasladado al intervalo donde se encuentra la intensidad de \mathbf{x} . La ventaja de esto es que ya no es necesario calcular la convolución con el kernel \mathcal{K} para cada voxel $\mathbf{x} \in L$, sino que basta con centrar el kernel en el intervalo del impulso, lo cual hace al método más eficiente. A continuación se detalla el algoritmo.

Algorithm 4.2.1: Histograma Local Ponderado Suavizado

Input: Un campo muestral t , un kernel Gaussiano W_j de varianza σ_j^2 , un kernel Gaussiano \mathcal{K} de varianza σ_l^2 y un número de intervalos B .

Output: El conjunto de histogramas locales ponderados para cada $\mathbf{x} \in L$.

- 1 Generar el cubo δ con (4.2.1).
- 2 Aplicar suavizado a cada $\delta_{\mathbf{x}}$, esto es, $\hat{\delta} = \delta \otimes \mathcal{K}$, usando (4.2.12).
- 3 Suavizar cada $\tilde{\delta}_i$ con kernel W_j usando

$$\tilde{h}_{i, W_j}(\mathbf{x}) = \sum_{\mathbf{y} \in W_{j\mathbf{x}}} \tilde{\delta}_i(\mathbf{y}) W_j(\mathbf{x} - \mathbf{y}). \quad (4.2.13)$$

- 4 **return** $\{\tilde{h}_{\mathbf{x}, W_j}(i) : \mathbf{x} \in L \text{ y } 1 \leq i \leq B\}$, esto es, el cubo \tilde{h}_{W_j} .
-

4.2.1. Método de Cuantil Local

La distribución acumulada local en el voxel \mathbf{x} , $\tilde{H}_W(\mathbf{x}, i)$, asociada al histograma local ponderado suavizado $\tilde{h}_W(\mathbf{x}, i)$, la podemos obtener con

$$\tilde{H}_W(\mathbf{x}, i) = d_0 \sum_{k=1}^i \tilde{h}_W(\mathbf{x}, k), \quad (4.2.14)$$

donde $d_0 = \frac{1}{\sum_{k=1}^B \tilde{h}_W(\mathbf{x}, k)}$. Usando la ecuación (4.2.14), podemos obtener fácilmente el cuantil de nivel p mediante la siguiente función

$$\tilde{H}_{\mathbf{x}, W}^{-1}(p) = \min_{\{i: p \leq \tilde{H}_{\mathbf{x}, W}(i)\}} \{b_1, \dots, b_B\}. \quad (4.2.15)$$

A continuación se describe el algoritmo del cuantil local.

Algorithm 4.2.2: Cuantil Local

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de Gaussianas, un kernel Gaussiano \mathcal{K} de varianza σ_I^2 , y el nivel p para el cuantil.

Output: Un conjunto de umbrales $\{\eta_{W_j}^*\}$.

1 Generar muestras de campos t usando M_0 .

2 **foreach** *campo muestral* t **y cada** $W_j \in \mathcal{W}$ **do**

3 Obtener el histograma $\tilde{h}_{W_j}(\mathbf{x}, i)$ con el algoritmo 4.2.1.

4 Para cada $\mathbf{x} \in L$ calcular $\theta_{W_j}(\mathbf{x}) = \tilde{H}_{\mathbf{x}, W_j}^{-1}(p)$ usando (4.2.15).

5 Calcular la variable aleatoria

$$\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x}) \quad (4.2.16)$$

6 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} para cada $W_j \in \mathcal{W}$.

7 **foreach** *campo muestral* t **do**

8 **foreach** $W_j \in \mathcal{W}$ **y cada** $\mathbf{x} \in L$ **do**

9 Repetir proceso de las líneas 3 y 4.

10 Calcular $g(\mathbf{x})$ con la ecuación (3.1.23).

11 Calcular g con la ecuación (3.1.24).

12 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .

13 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.

14 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales $\eta_{W_j}^*$ usando la ecuación (3.1.26).

15 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.

Algorithm 4.2.3: Cuantil Local (Segunda etapa)

Input: Un campo de prueba t , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de Gaussianas, un conjunto de umbrales de detección correspondiente $\{\eta_{W_j}^*\}$, y un kernel Gaussiano \mathcal{K} de varianza σ_I^2 .

Output: Una estimación de RA \hat{A} en t .

1 Repetir las líneas 3 y 4 del algoritmo 4.2.2.

2 Calcular \tilde{A} con (4.2.17).

3 Calcular el estimador $\hat{A} = \bigcup_{j=1}^N \tilde{A}_{W_j}$.

4 **return** \hat{A} .

donde

$$\tilde{A}_{W_j}(\mathbf{x}) = \begin{cases} 1, & \text{si } \theta_{W_j}(\mathbf{x}) > \eta_{W_j}^*; \\ 0, & \text{en otro caso.} \end{cases} \quad (4.2.17)$$

Para el caso particular en el que $p = 0.5$ llamaremos al procedimiento descrito con los algoritmos 4.2.2 y 4.2.3 Método de Mediana Local. En el capítulo 5 de experimentos y resultados sólo se realizan pruebas usando dicho método.

4.2.2. Método de Moda Local

Una vez que se tiene el histograma local ponderado suavizado con el algoritmo 4.2.1, es sencillo calcular la moda local en \mathbf{x} asociada a W_j . Para hallarla, únicamente es necesario localizar el intervalo b_k de mayor frecuencia. El método de moda local se obtiene a partir del procedimiento del cuantil local, basta con cambiar la línea 4 del algoritmo 4.2.2 por lo siguiente:

Para cada $\mathbf{x} \in L$ hallar k tal que

$$\tilde{h}_{\mathbf{x}, W_j}(k) = \max_{1 \leq i \leq B} \left\{ \tilde{h}_{\mathbf{x}, W_j}(i) \right\} \quad (4.2.18)$$

y hacer

$$\theta_{W_j}(\mathbf{x}) = b_k. \quad (4.2.19)$$

El algoritmo de la segunda parte de moda local permanece igual a 4.2.3, teniendo en cuenta que en este caso $\theta_{W_j}(\mathbf{x})$ es como en (4.2.19).

4.2.3. Método de Filtro Bilateral

A continuación se plantean tres métodos que están basados en la idea del llamado *Filtro Bilateral* y, al igual que en los casos anteriores se utiliza el histograma local ponderado suavizado. Antes de definir los algoritmo daremos una breve explicación de algunos conceptos que serán de utilidad para su comprensión.

Aplicar un *filtro* a una imagen digital es una de las operaciones fundamentales en procesamiento de imágenes y visión computacional. Filtrar una imagen consiste en asignar a cada píxel \mathbf{x} en la imagen de salida I_s el valor de una función que depende de una vecindad del píxel correspondiente en la imagen original I_e . Si la función que se utiliza es una combinación lineal de los valores de la imagen de entrada con coeficientes positivos (promedio pesado, por ejemplo, pesos dados por una Gaussiana), entonces el resultado será una imagen más suave que la original. Si los valores de salida se calculan a partir de una vecindad de píxeles relativamente grande, la imagen de salida podría resultar “demasiado” borrosa.

El *filtro bilateral* fue desarrollado en [29, Tomasi and Manduchi, (1998)], y es un filtro no

lineal de suavizado que preserva los bordes fuertes y reduce el ruido³. La salida de este filtro es un promedio pesado de la entrada. Para asignar el peso a un píxel vecino \mathbf{y} de un píxel dado \mathbf{x} , se utiliza un kernel Gaussiano espacial w (al decir espacial nos referimos a que depende de la distancia que hay entre \mathbf{y} y \mathbf{x}) y una función \mathcal{K}_R en el dominio de la intensidad, que disminuye el peso entre más grande sea la diferencia de intensidades de los píxeles. Esto quiere decir que el peso, además de depender de la distancia euclidiana $\|\mathbf{y} - \mathbf{x}\|$, también depende de $|I_e(\mathbf{y}) - I_e(\mathbf{x})|$. El filtro bilateral Gaussiano en un píxel \mathbf{x} se define como:

$$f(\mathbf{x}) = c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{y} - \mathbf{x}) \mathcal{K}_R(I_e(\mathbf{y}) - I_e(\mathbf{x})) I_e(\mathbf{y}), \quad (4.2.20)$$

donde

$$w(\mathbf{x}) = e^{-\frac{1}{2} \frac{\|\mathbf{x}\|^2}{\sigma^2}}, \quad (4.2.21)$$

para $\mathbf{x} = (x_1, x_2)$, $W_{\mathbf{x}}$ es el conjunto de elementos que pertenecen al kernel w centrado en \mathbf{x} , la función

$$\mathcal{K}_R(x) = e^{-\frac{1}{2} \frac{x^2}{\sigma_R^2}} \quad (4.2.22)$$

para x escalar, y c es un término de normalización tal que

$$c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{y} - \mathbf{x}) \mathcal{K}_R(I_e(\mathbf{y}) - I_e(\mathbf{x})) = 1,$$

Para adecuar la ecuación (4.2.20) a la notación e ideas que hemos venido manejando, suponga que se tiene un modelo M_0 para generar campos aleatorios discretos t bajo la hipótesis nula (en este caso la imagen de entrada es el campo t), y que nuevamente hemos dividido el rango de intensidades en B intervalos. Denotaremos con i_x el intervalo tal que $t(\mathbf{x}) \in [l_{i_x}, l_{i_x+1})$ y con i_y el intervalo tal que $t(\mathbf{y}) \in [l_{i_y}, l_{i_y+1})$. Entonces

$$\begin{aligned} f(\mathbf{x}) &= c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{y} - \mathbf{x}) \mathcal{K}_R(I_e(\mathbf{y}) - I_e(\mathbf{x})) I_e(\mathbf{y}) \\ &= c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{y} - \mathbf{x}) \mathcal{K}_R(t(\mathbf{y}) - t(\mathbf{x})) t(\mathbf{y}) \\ &\approx c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{y} - \mathbf{x}) \mathcal{K}_R(b_{i_y} - b_{i_x}) b_{i_y} \\ &= c \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{x} - \mathbf{y}) \mathcal{K}_R(b_{i_x} - b_{i_y}) b_{i_y} \end{aligned} \quad (4.2.23)$$

pues $\|\mathbf{x} - \mathbf{y}\| = \|\mathbf{y} - \mathbf{x}\|$ y $(x - y)^2 = (y - x)^2$.

³Reducción de ruido se entiende como disminuir la diferencia de intensidades que hay entre los píxeles y sus vecinos.

Por otro lado, si para cada $\mathbf{y} \in W_{\mathbf{x}}$ se tiene que $t(\mathbf{y}) \in [l_{i_y}, l_{i_y+1})$ entonces $\delta_{\mathbf{y}}(i)$ es un impulso en $i = i_y$ y por tanto

$$\sum_{i=1}^B \delta_{\mathbf{y}}(i) \mathcal{K}_R(b_k - b_i) b_i = \mathcal{K}_R(b_k - b_{i_y}) b_{i_y} \quad (4.2.24)$$

para cualquier $k \in \{1, \dots, B\}$.

Entonces, tomando $k = i_x$ en (4.2.24) y sustituyendo en (4.2.23) nos queda

$$\begin{aligned} f(\mathbf{x}) &= c \sum_{\mathbf{y} \in W_{\mathbf{x}}} \left[w(\mathbf{x} - \mathbf{y}) \left(\sum_{i=1}^B \delta_{\mathbf{y}}(i) \mathcal{K}_R(b_{i_x} - b_i) b_i \right) \right] \\ &= c \sum_{\mathbf{y} \in W_{\mathbf{x}}} \left(\sum_{i=1}^B w(\mathbf{x} - \mathbf{y}) \delta_{\mathbf{y}}(i) \mathcal{K}_R(b_{i_x} - b_i) b_i \right) \\ &= c \sum_{i=1}^B \sum_{\mathbf{y} \in W_{\mathbf{x}}} w(\mathbf{x} - \mathbf{y}) \delta_{\mathbf{y}}(i) \mathcal{K}_R(b_{i_x} - b_i) b_i \\ &= c \sum_{i=1}^B \left[\left(\sum_{\mathbf{y} \in W_{\mathbf{x}}} \delta_{\mathbf{y}}(i) w(\mathbf{x} - \mathbf{y}) \right) \mathcal{K}_R(b_{i_x} - b_i) b_i \right] \\ &= c \sum_{i=1}^B \left[\left(\sum_{\mathbf{y} \in W_{\mathbf{x}}} \delta_i(\mathbf{y}) w(\mathbf{x} - \mathbf{y}) \right) \mathcal{K}_R(b_{i_x} - b_i) b_i \right] \\ &= c \sum_{i=1}^B \left[\left(\hat{h}_{\mathbf{x}, W}(i) \right) \mathcal{K}_R(b_{i_x} - b_i) b_i \right] \\ &= c \sum_{i=1}^B \mathbf{h}_{\mathbf{x}, W}(i) b_i \end{aligned} \quad (4.2.25)$$

donde

$$\mathbf{h}_{\mathbf{x}, W}(i) = \hat{h}_{\mathbf{x}, W}(i) \mathcal{K}_R(b_{i_x} - b_i) \quad (4.2.26)$$

y

$$c = \frac{1}{\sum_{i=1}^B \mathbf{h}_{\mathbf{x}, W}(i)}$$

es término de normalización. Luego, si para cada píxel $\mathbf{x} \in L$ si multiplicamos el histograma local ponderado por la función \mathcal{K}_R centrada en el intervalo i_x (donde i_x es tal que $t(\mathbf{x}) \in [l_{i_x}, l_{i_x+1})$) obtenemos $\mathbf{h}_{\mathbf{x}, W}$.

Entonces, podemos calcular el filtro bilateral en \mathbf{x} simplemente calculando la media local con $\mathbf{h}_{\mathbf{x}, W}$. Basados en la idea anterior, realizamos algunas variantes de este método. En lugar de utilizar $\hat{h}_{\mathbf{x}, W}$, emplearemos el histograma local ponderado suavizado $\tilde{h}_{\mathbf{x}, W}$.

Así, el primer método consiste en calcular la media local con

$$\mu(x) = c \sum_{i=1}^B \tilde{h}_{\mathbf{x}, W}(i) b_i, \quad (4.2.27)$$

donde

$$\bar{h}_{\mathbf{x},W}(i) = \tilde{h}_{\mathbf{x},W}(i)\mathcal{K}_R(b_{i_x} - b_i), \quad (4.2.28)$$

y

$$c = \frac{1}{\sum_{i=1}^B \bar{h}_{\mathbf{x},W}(i)}.$$

A continuación se enuncian los algoritmos 4.2.4 y 4.2.5 que describen el procedimiento anterior.

Algorithm 4.2.4: Filtro Bilateral Local

Input: El modelo estadístico M_0 para generar campos aleatorios bajo H_0 , un nivel de significancia α , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de Gaussianas, un kernel Gaussiano \mathcal{K}_I de varianza σ_I^2 y una función Gaussiana \mathcal{K}_R de varianza σ_R^2 .

Output: Un conjunto de umbrales $\{\eta_{W_j}^*\}$.

1 Generar muestras de campos t usando M_0 .

2 **foreach** *campo muestral* t y *cada* $W_j \in \mathcal{W}$ **do**

3 Calcular $\hat{h}_{W_j}(\mathbf{x}, i)$ con algoritmo 4.2.1.

4 Para cada $\mathbf{x} \in L$ hallar i_x tal que $t(\mathbf{x}) \in [l_{i_x}, l_{i_x+1})$.

5 Calcular $\bar{h}_{\mathbf{x}, W}(i)$ (4.2.28).

6 Para cada $\mathbf{x} \in L$ hallar $\theta_{W_j}(\mathbf{x}) = \mu(\mathbf{x})$ con (4.2.27).

7 Calcular la variable aleatoria

$$\theta_{W_j} = \max_{\mathbf{x} \in L} \theta_{W_j}(\mathbf{x}). \quad (4.2.29)$$

8 Encontrar la distribución empírica $F_{\theta_{W_j}}$ de los valores θ_{W_j} para cada $W_j \in \mathcal{W}$.

9 **foreach** *campo muestral* t **do**

10 **foreach** $W_j \in \mathcal{W}$ y *cada* $\mathbf{x} \in L$ **do**

11 Repetir proceso de la línea 3 a la 6 para obtener $\theta_{W_j}(\mathbf{x})$.

12 Calcular $g(\mathbf{x})$ con la ecuación (3.1.23).

13 Calcular g con la ecuación (3.1.24)

14 Encontrar la distribución empírica $F_g(\cdot)$ de los valores g .

15 Para un nivel de significancia α dado, encontrar g^* tal que $F_g(g^*) = 1 - \alpha$.

16 Para cada $W_j \in \mathcal{W}$ encontrar los umbrales $\eta_{W_j}^*$ usando la ecuación (3.1.26).

17 **return** $\{\eta_{W_1}^*, \dots, \eta_{W_N}^*\}$.

Algorithm 4.2.5: Filtro Bilateral Local (Segunda etapa)

Input: Un campo de prueba t , una familia $\mathcal{W} = \{W_1, \dots, W_N\}$ de Gaussianas, un conjunto de umbrales de detección correspondiente $\{\eta_{W_j}^*\}$, un kernel Gaussiano \mathcal{K}_I de varianza σ_I^2 , y una función Gaussiana \mathcal{K} de varianza σ_R^2 .

Output: Una estimación de RA \hat{A} en t .

1 Repetir las líneas 3-6 del algoritmo 4.2.4.

2 Calcular $\tilde{A}_{W_j}(x)$ como en (4.2.17) de algoritmo 4.2.3 pero $\theta_{W_j}(\mathbf{x})$ como en la línea 6 de algoritmo 4.2.4.

3 Calcular el estimador $\hat{A} = \bigcup_{j=1}^N \tilde{A}_{W_j}$.

4 **return** \hat{A} .

El segundo método consiste en calcular la mediana local en \mathbf{x} con $\bar{h}_{\mathbf{x},W}(i)$. Para ello cambiamos la línea 6 del algoritmo 4.2.4 por:

Para cada $x \in L$ calcular

$$\theta_{W_j}(\mathbf{x}) = \bar{H}_{x,W_j}^{-1}(0.5), \quad (4.2.30)$$

donde

$$\bar{H}_{x,W_j}^{-1}(p) = \min_{i:p \leq \bar{H}_{x,W_j}(i)} \{b_1, \dots, b_B\}, \quad (4.2.31)$$

con $p = 0.5$ y

$$\bar{H}_{x,W_j}(i) = \sum_{k=1}^i \bar{h}_{x,W_j}(k). \quad (4.2.32)$$

En el tercer y último método, encontramos la moda local en \mathbf{x} usando $\bar{h}_{\mathbf{x},W}(i)$. Con este fin, modificaremos el paso 6 del algoritmo 4.2.4 por:

Para cada $x \in L$ hallar k tal que

$$\bar{h}_{\mathbf{x},W_j}(k) = \max_{1 \leq i \leq B} \{\bar{h}_{\mathbf{x},W_j}(i)\} \quad (4.2.33)$$

y hacer

$$\theta_{W_j}(x) = b_k. \quad (4.2.34)$$

Experimentos y Resultados

Para evaluar el desempeño de los métodos descritos en los capítulos 3 y 4 se realizaron dos tipos de experimentos: simulación de campos Gaussianos y datos sintéticos fMRI. En general se hace una comparación entre los resultados obtenidos usando FWER y MBHT adaptable con el resto de los métodos.

Con el fin de determinar y comparar la calidad de las estimaciones de las regiones de activación obtenidas con cada uno de los distintos métodos, empleamos como medida el índice de Tanimoto (también conocido como índice Jaccard), definido por

$$I_T = \frac{|A \cap \hat{A}|}{|A \cup \hat{A}|}, \quad (5.0.1)$$

recordando que A es la región de activación y \hat{A} es la región de activación estimada. Es importante observar que si $I_T = 1$ entonces $\hat{A} = A$, mientras que si $I_T = 0$ implica que A y \hat{A} son conjuntos disjuntos, es decir $A \cap \hat{A} = \emptyset$. El índice de Tanimoto es afectado tanto por la sensibilidad como por la selectividad, por lo que es importante verificar que los errores de falsos positivos (selectividad) estén bajo control.

En las siguientes secciones se describe en que consiste cada experimento, comenzando con la simulación de campos Gaussianos.

5.1. Simulación de Campos Gaussianos

Para realizar esta simulación, la señal t se generó con el siguiente modelo

$$t = f + \mathcal{N}, \quad (5.1.1)$$

donde \mathcal{N} es un campo de variables aleatorias que se distribuyen como una Gaussiana de media cero y varianza igual a la unidad. El campo f se estableció igual a cero para cada píxel que no estuviera contenido en la región de activación, esto es, $f(x) = 0$ para cada $x \in L$ tal que $x \notin A$.

La región de activación A consiste de dos cuadrados con nivel de activación a , de 9 y 5 píxeles de extensión respectivamente. A dicha región de activación, se le aplica un suavizado con un kernel Gaussiano de varianza $\sigma = 0.5$, la RA resultante se trunca para valores de intensidad menores a $0.5a$. Para cada nivel de activación a se generaron 50 imágenes de prueba con el modelo (5.1.1) (bajo el supuesto de que la región de activación $A \neq \emptyset$) y para cada imagen se calculo el índice de Tanimoto. Se obtuvo un promedio del índice en cada nivel de activación a y con estos promedios se construyó la curva de Tanimoto de cada método.

Para la simulación se consideran dos casos. En el primero \mathcal{N} es un campo de ruido blanco, lo que quiere decir que no hay correlación espacial entre los píxeles (ver imagen 5.1.1.b). En el segundo se introduce correlación espacial al convolucionar la imagen de ruido (\mathcal{N}) con un kernel Gaussiano de parámetro $\sigma = 1$ píxel como aparece en la figura 5.1.1.c).

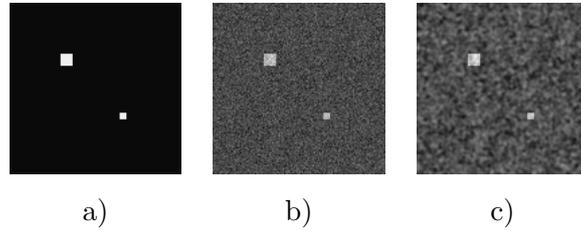


Fig. 5.1.1: a) Imagen de región de activación A con SNR=15; b) Imagen de activación, $a = 6.0$, con ruido blanco aditivo (no hay correlación espacial entre píxeles); c) Imagen de región de activación, $a = 1.8$, con ruido correlacionado espacialmente (el ruido se convoluciona con un kernel Gaussiano de $\sigma = 1$ píxel).

El tamaño de la rejilla L que usamos es de $m \times m$ píxeles, con $m = 128$. Los umbrales estimados, $\theta_{W_j}^*$ y $\eta_{W_j}^*$, se obtuvieron con un nivel de confianza $\alpha = 0.05$ en cada uno de los casos donde fue necesario calcularlos. Se utilizaron conjuntos de 1000 imágenes con rejilla de píxeles L bajo el supuesto de la hipótesis nula (es decir que, $f(x) = 0$ para toda $x \in L$), para calcular cada una de las distribuciones empíricas F_{θ_0} , $F_{\theta_{W_j}}$ y F_g .

En todas las figuras la curva del índice de Tanimoto del método MBHT Adaptable se obtuvo usando una familia de $N = 5$ elementos estructurales de radios $\{0, 1, 2, 3, 4\}$.

En la sección 4.2, que describe los métodos basados en el *histograma local ponderado suavizado*, quedó pendiente la pregunta de cómo elegir adecuadamente el valor de σ_I para el kernel de suavizado \mathcal{K}_x . Para hallar este valor, primero se obtuvieron los umbrales $\eta_{W_j}^*$ de cada método para distintos valores de σ_I , los cuales tomamos del intervalo $[0, 2]$ con incrementos de 0.25. Con estos umbrales se obtuvo un promedio del índice de Tanimoto en cada nivel de activación a para construir la curva de Tanimoto correspondiente. En cada caso, se eligió el valor de σ_I , para el cual se obtuvo el mejor desempeño. De hecho, con este procedimiento no sólo obtuvimos el valor de σ_I sino también la familia de Gaussianas apropiada. Para la obtención de los umbrales $\eta_{W_j}^*$ se consideraron familias de Gaussianas de $N = 6, 9, 13$ elementos, con varianzas de $\sigma_1 =$

$0.0, \sigma_2 = 0.25, \dots, \sigma_N = 0.25(N - 1)$. Así como elementos individuales en lugar de familias de Gaussianas, es decir, probamos cada método con una sola Gaussianas para distintos valores de $\sigma_1 = 0.0, 0.25, \dots, 3.0$. De esta manera, en cada caso, seleccionamos a σ_I y la familia de Gaussianas óptimas.

En primer lugar se muestran las gráficas de las curvas del índice de Tanimoto, bajo la consideración de que no hay correlación espacial (ruido blanco), para lo cual se tomaron 20 niveles de activación $a = 0.0, 0.4, 0.8, \dots, 7.6$.

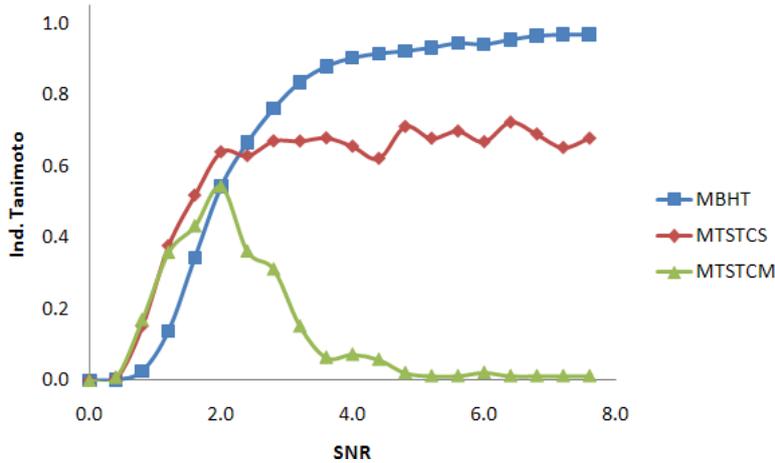


Fig. 5.1.2: Comparación entre las curvas del índice de Tanimoto obtenidas con los métodos: Morphology-Based Hypothesis Testing Adaptable (MBHT), Multi-Threshold Supra Threshold Cluster Size (MTSTCS) y Multi-Threshold Supra Threshold Cluster Mass (MTSTCM) para ruido blanco.

En la figura 5.1.2 se usaron $N = 40$ umbrales de intensidad $\{0.0, 0.1, 0.2, \dots, 3.9\}$ para encontrar los umbrales de área y masa con los métodos MTSTCS y MTSTCM respectivamente. En general se observa que MBHT tiene un mejor desempeño para la detección de RA. Como se puede observar en la figura 5.1.11, los métodos basados en clusters conectados pueden resultar en estructuras porosas. Por ello, MTSTCS y MTSTCM presentan cierto grado de inestabilidad cuando la diferencia entre la señal y el ruido es grande.

En la figura 5.1.3 el método *TFCE* se corrió empleando la misma familia de 40 umbrales de intensidad que en la figura 5.1.2, mientras que para el método MBHT-TFCE se usó una familia \mathcal{W} de 6 SE con radios $\{0, 1, 2, 3, 4, 5\}$. Aquí observamos que el comportamiento de los tres métodos es similar, aunque el MBHT resulta ser un poco más sensible para detectar la región de activación cuando la relación señal ruido es relativamente pequeña en comparación con el procedimiento TFCE.

En la figura 5.1.4, para el método basado en estadísticos de orden n utilizamos una familia \mathcal{W} de $N = 5$ SE con radios $\{0, 1, 2, 3, 4\}$. Se probó con distintos valores de n , sin embargo los mejores resultados se obtuvieron con $n = 10$. Para valores más grandes las curvas mostraron

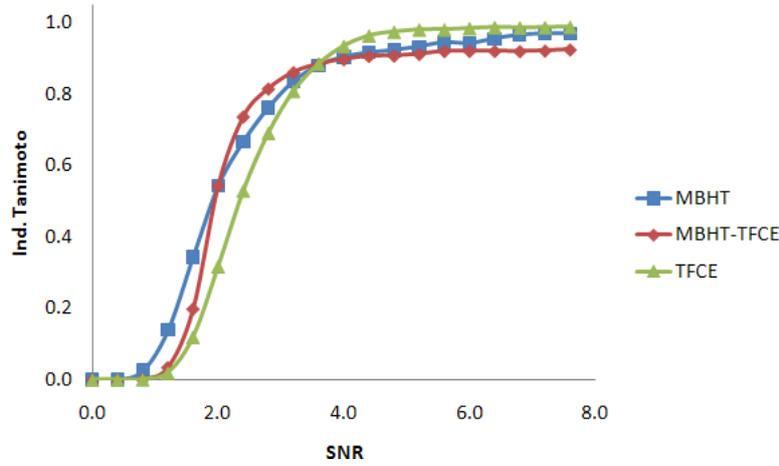


Fig. 5.1.3: Curvas de Tanimoto de: Morphology-Based Hypothesis Testing Adaptable (MBHT), Threshold Free Cluster Enhancement (TFCE) y la combinación de ambos (MBHT-TFCE) para ruido blanco.

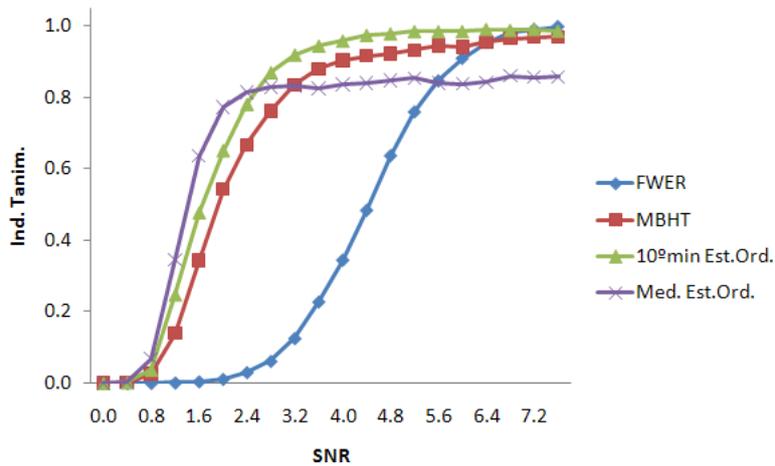


Fig. 5.1.4: Curvas índice de Tanimoto de: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT), el estadístico de orden 10 y la mediana con estadísticos de orden.

una ligera mejoría del método en la sensibilidad para detectar RA cuando la relación señal ruido es pequeña, pero tienden a sobrestimar la RA cuando la relación señal ruido es grande. Esto puede apreciarse con la curva que se obtuvo de la mediana (local) con estadísticos de orden. La gráfica muestra que el procedimiento del 10^o mín tiene el mejor desempeño.

Para los métodos de mediana y moda local de la figura 5.1.5 fue necesario discretizar el rango de $[-4, 5]$ en $B = 100$ intervalos. Además se usó una familia de $N = 6$ Gaussianas de desviaciones estándar $\{0.0, 0.25, 0.5, 1.0, 1.25, 1.5\}$ y $N = 9$ con $\sigma = 0.0, 0.25, \dots, 2.0$ para la mediana y moda local respectivamente. Se probó con varios valores para σ_I para el kernel Gaussiano \mathcal{K} y los mejores resultados ocurrieron en $\sigma_I = 1.25$ para ambos métodos. La gráfica muestra que tanto el método de la mediana como el de la moda local superan al método MBHT, e incluso se

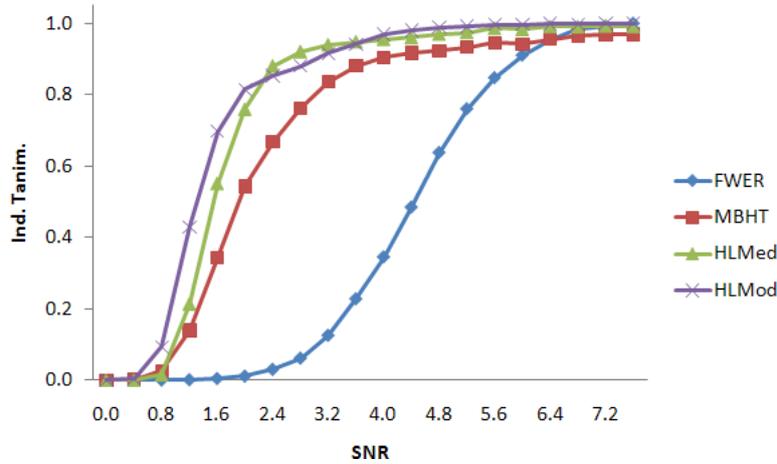


Fig. 5.1.5: Curvas del índice de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT), Mediana Local, y Moda Local para ruido no correlacionado.

distingue un mejor desempeño con la moda local.

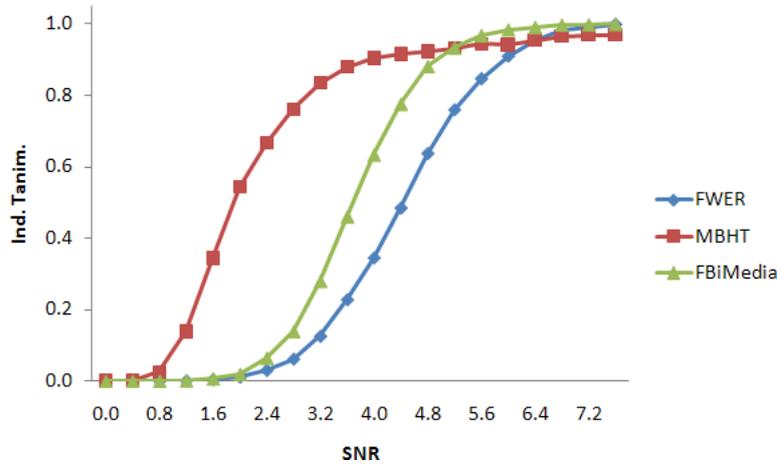


Fig. 5.1.6: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Media Local con 4.2.27 (variante de filtro bilateral) para ruido no correlacionado.

Para la media local de la figura 5.1.6 también dividimos el rango de intensidad $[-4, 5]$ en $B = 100$ intervalos. Se usó el histograma local ponderado suavizado, y una familia de $N = 9$ Gaussianas con $\sigma = 0.0, 0.25, \dots, 2.0$. Para la función Gaussiana \mathcal{K}_R se tomó $\sigma_R = 1$ y, después de probar con varios valores para σ_I , se eligió $\sigma_I = 1.25$.

Usando la idea de Filtro Bilateral, en la figura 5.1.7 calculamos la mediana local utilizando \bar{h}_{x, W_j} , ya que esta considera no solo la distancia espacial sino también la diferencia entre las intensidades de píxeles vecinos. Para ello se usó una familia de $N = 9$ Gaussianas como en la gráfica 5.1.6, un kernel Gaussiano \mathcal{K} de $\sigma_I = 1.0$, $\sigma_R = 1.0$ y los mismos $B = 100$ intervalos.

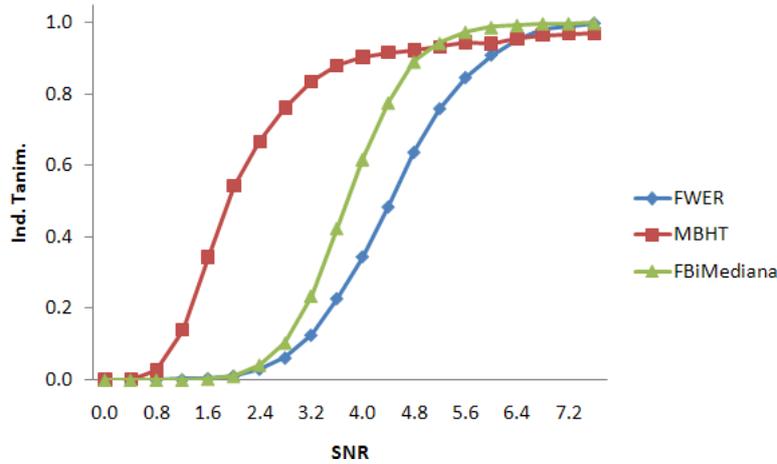


Fig. 5.1.7: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Mediana Local con 4.2.30 (variante de filtro bilateral) para ruido no correlacionado.

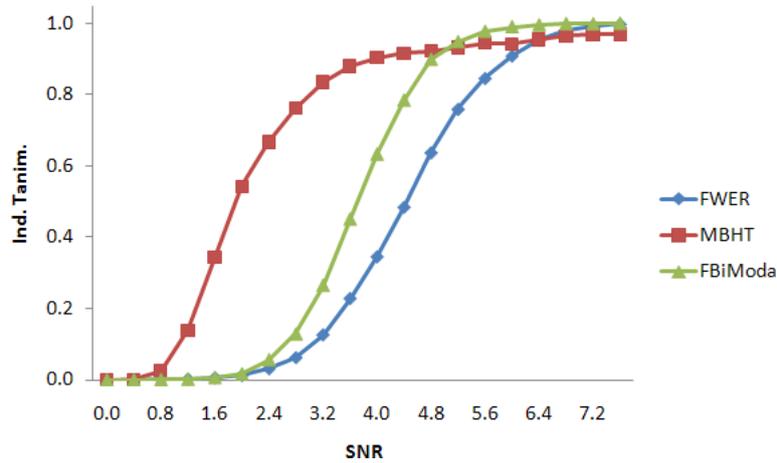


Fig. 5.1.8: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Moda Local con 4.2.34 (variante de filtro bilateral) para ruido no correlacionado.

Para calcular la moda local en la figura 5.1.8 también se obtuvo \bar{h}_{x,W_j} para cada $x \in L$. El mejor resultado que conseguimos fue con una familia de 9 Gaussianas con $\sigma = 0.0, 0.25, \dots, 2.0$, $\sigma_I = 1.25$, $\sigma_R = 1.0$ y la misma división de $B = 100$ intervalos.

Como puede verse en las figuras 5.1.6, 5.1.7 y 5.1.8 los métodos implementados bajo la idea de filtro bilateral son muy similares para ruido no correlacionado. Es claro ninguno de estos métodos fue mejor que el MBHT. Esto es un resultado interesante y contrario a lo que se esperaba.

La figura 5.1.9 muestra las gráficas del índice de Tanimoto de los métodos que mostraron un mejor desempeño para ruido no correlacionado. La moda local, obtenida con el histograma local ponderado suavizado, muestra una mayor sensibilidad para detectar RA con niveles de

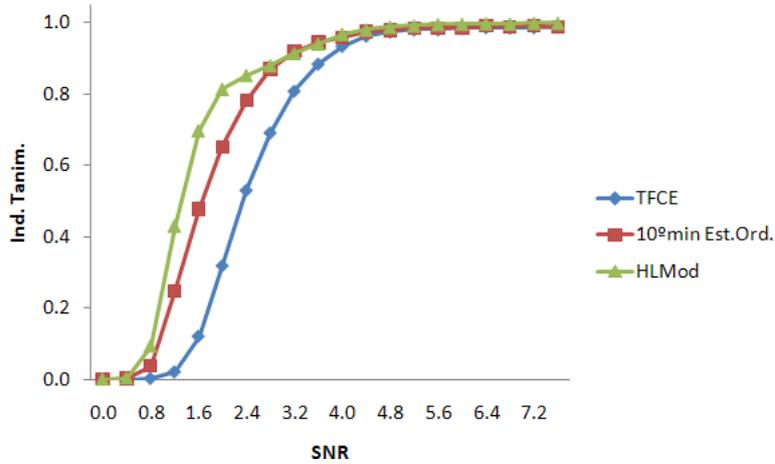


Fig. 5.1.9: Curvas de Tanimoto obtenidas con Threshold Free Cluster Enhancement (TFCE), el estadístico de orden 10 (10º mín) y Moda Local con histograma local ponderado (HLMod).

activación moderado en comparación con el resto.

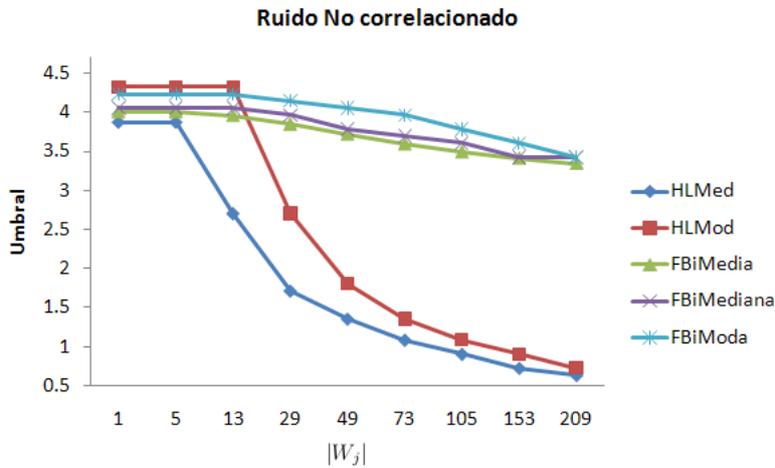


Fig. 5.1.10: Gráfica del valor del umbral $\eta_{W_j}^*$ contra el número de elementos que contiene el kernel w_j , para cada uno de los métodos basados en el histograma local ponderado suavizado.

La figura 5.1.10 muestra los umbrales $\eta_{W_j}^*$ de intensidad contra el número de elementos contenidos en el kernel w_j , para cada uno de los métodos que utilizan $\tilde{h}_{x,W}(i)$. Se usó una familia \mathcal{W} de $N = 9$ Gaussianas de desviación estándar $\sigma = 0.0, 0.25, \dots, 2.0$. Observe cómo las gráficas de la moda y mediana local decrecen rápidamente al incrementar el tamaño del kernel. Esto no sucede en los métodos basados en filtro bilateral y se debe principalmente a que el filtro bilateral es sensible a los datos atípicos.

Para darnos una idea de cuánto tardan los métodos, en la tabla 5.1.1 se muestra el tiempo en segundos para calcular los umbrales $\eta_{W_j}^*$ y detectar la región de activación en una imagen

Tiempo (en segundos)		
Métodos	Umbrales	Detec. Act.
MBHT	126.81431	0.04324
10 ^o mín Est.Ord.	738.47280	0.14855
HLMMod	6465.83074	3.21001
FBiMedia	7581.29169	3.77515

Tabla 5.1.1: En la primer columna se muestra el método seguido por los tiempos aproximados (en segundos) para calcular los umbrales $\eta_{W_j^*}$ (Umbrales) y detectar la región de activación en una imagen de prueba (Detec. Act.).

de prueba. Los tiempos se obtuvieron generando campos de ruido Gaussiano no correlacionado. Para el MBHT se usó una familia de cinco SE de radios $r = 0, 1, 2, 3, 4$; con 10^o mín se usó la misma familia que con MBHT; para Moda Local con \tilde{h}_W (histograma local ponderado suavizado) se usó una familia de nueve Gaussianas con $\sigma = 0.0, 0.25, \dots, 2.0$ y $\sigma_I = 1.25$; para la Media Local con \bar{h}_W (variante de filtro bilateral) se usaron los mismos parámetros que con la Moda Local y $\sigma_R = 1.0$.

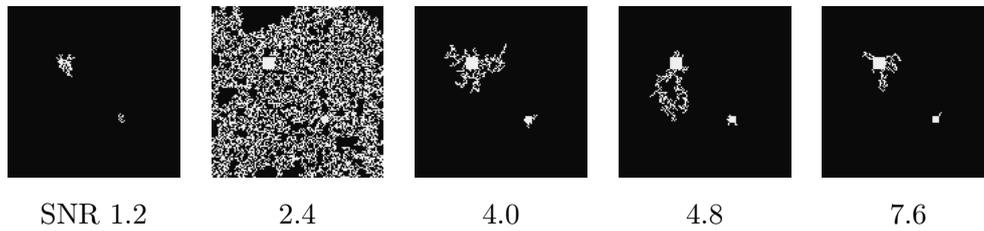


Fig. 5.1.11: Región de activación estimada con el método MTSTCS con diferentes valores de SNR, para ruido blanco (no correlacionado).

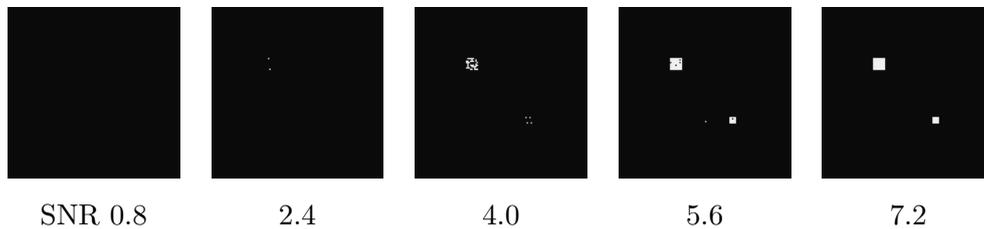


Fig. 5.1.12: Región de activación estimada con el método FWER con diferentes valores de SNR, para ruido blanco.

En las figuras 5.1.11 a 5.1.19 se muestran las regiones de activación estimadas que se obtuvieron con los métodos MTSTCS, FWER, TFCE, MBHT, MBHT-TFCE, 10^o mín, Mediana con Estadísticos Orden, Moda Local (con $\tilde{h}_{\mathbf{x},W}$) y Filtro Bilateral (media local con $\bar{h}_{\mathbf{x},W}$). Se puede

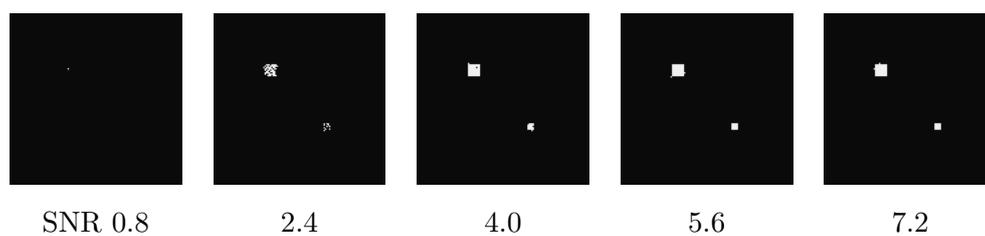


Fig. 5.1.13: Región de activación estimada con el método TFCE con diferentes valores de SNR, para ruido no correlacionado.

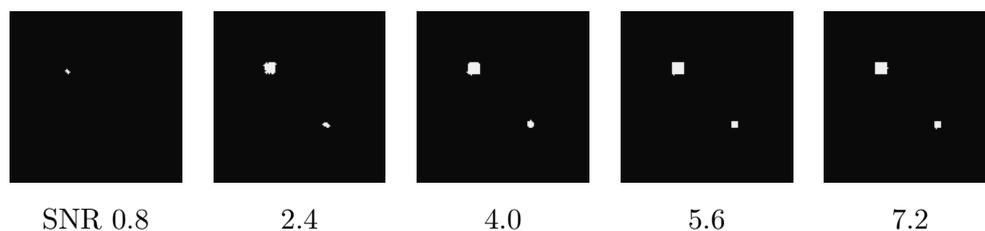


Fig. 5.1.14: Estimación de la región de activación con el método MBHT en distintos valores de SNR.

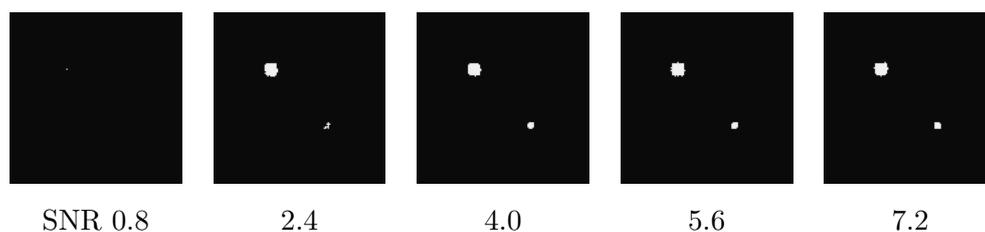


Fig. 5.1.15: Estimación de la región de activación con el método MBHT-TFCE en distintos valores de SNR.

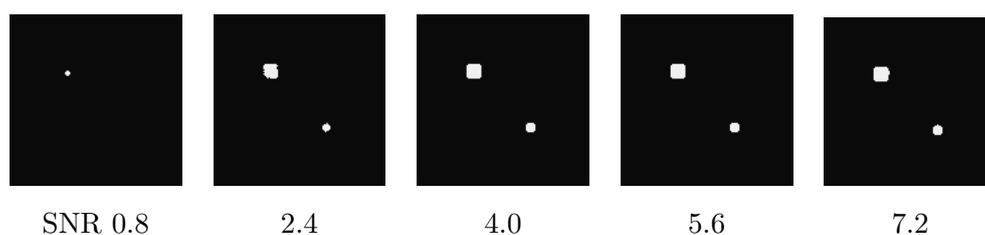


Fig. 5.1.16: Región de activación \hat{A} estimada con el 10^2 mín para diversos valores de SNR.

observar que con el método de la moda local, los bordes de la región de activación estimada quedan mejor definidos en comparación con el resto de los métodos; además los resultados no son estructuras porosas como los obtenidos con los métodos basados en STCC (ver figura 5.1.11).

Veamos ahora las curvas del índice de Tanimoto que obtenemos con estos mismos métodos, pero aplicados a imágenes de prueba con ruido correlacionado espacialmente. Para generar este tipo de imágenes se convoluciona el campo \mathcal{N} con un kernel Gaussiano con desviación estándar

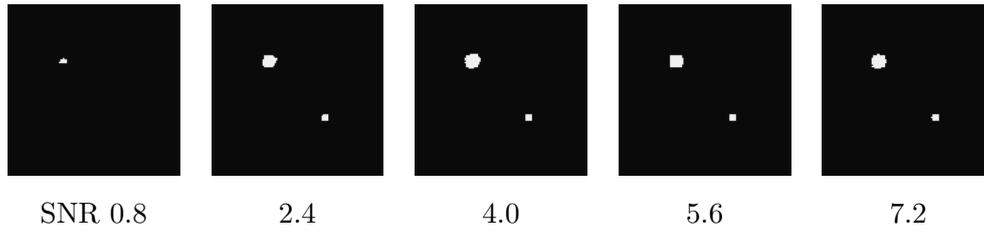


Fig. 5.1.17: Región de activación \hat{A} estimada con la Mediana Local de Estadísticos de Orden para diversos valores de SNR.

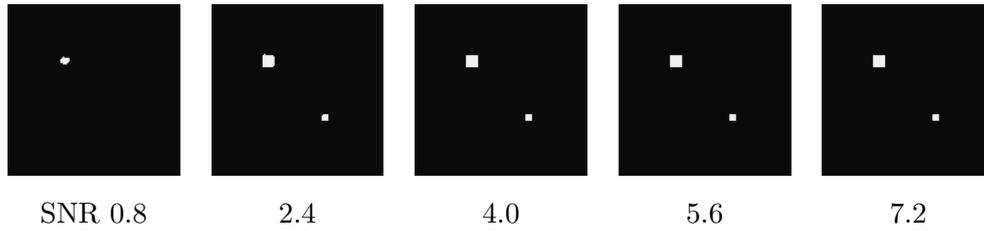


Fig. 5.1.18: Región de activación que obtuvimos como resultado con el método de la Moda Local con histograma local ponderado suavizado.

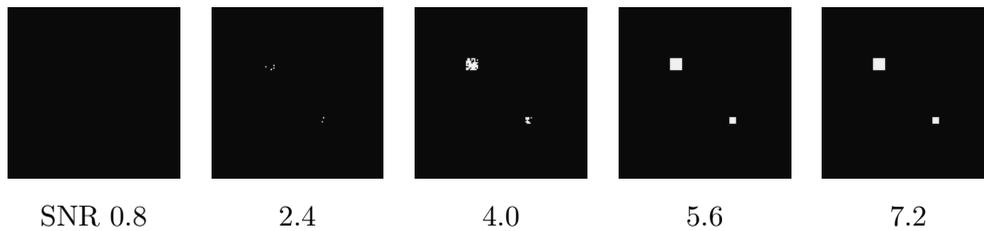


Fig. 5.1.19: Región de activación que obtuvimos como resultado con el método de Filtro Bilateral para Media Local.

$\sigma = 1$. Para estas pruebas se usaron 20 niveles de activación $a = 0.0, 0.12, \dots, 2.28$.

Para los métodos MTSTCS y MTSTCM de la figura 5.1.20 se usaron $N = 40$ umbrales de intensidad $\{0.0, 0.1, \dots, 3.9\}$. Nuevamente observamos que ambos métodos siguen teniendo problemas para detectar la RA cuando la relación señal ruido es grande.

En el método TFCE de la figura 5.1.21 se usó la familia de umbrales $\{0.0, 0.1, \dots, 3.9\}$, y para el algoritmo MBHT-TFCE usamos una familia \mathcal{W} de 6 SE de radios $\{0, 1, 2, 3, 4, 5\}$. En este caso, se observa que el TFCE se comporta mejor que el MBHT y el MBHT-TFCE.

Para el estadístico de orden n y para la mediana de estadísticos de orden usamos una familia \mathcal{W} de $N = 5$ SE de radios $r = 0, 1, 2, 3, 4$. Aquí podemos observar que el comportamiento del 10^o mín es muy similar al MBHT; sin embargo, el comportamiento de la mediana local (con Estadísticos de Orden) es muy distinto al que se observa en la gráfica 5.1.4, de hecho tiene el mejor desempeño.

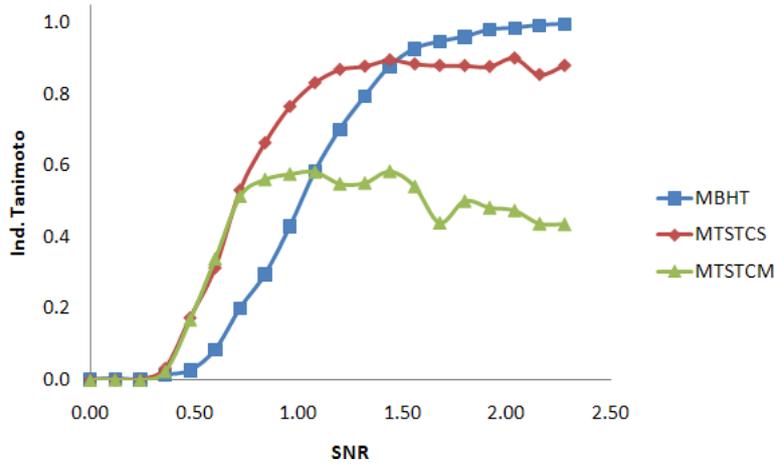


Fig. 5.1.20: Curvas del índice de Tanimoto obtenidas con los métodos: Morphology-Based Hypothesis Testing Adaptable (MBHT), Multi-Threshold Supra Threshold Cluster Size (MTSTCS) y Multi-Threshold Supra Threshold Cluster Mass (MTSTCM) para ruido correlacionado.

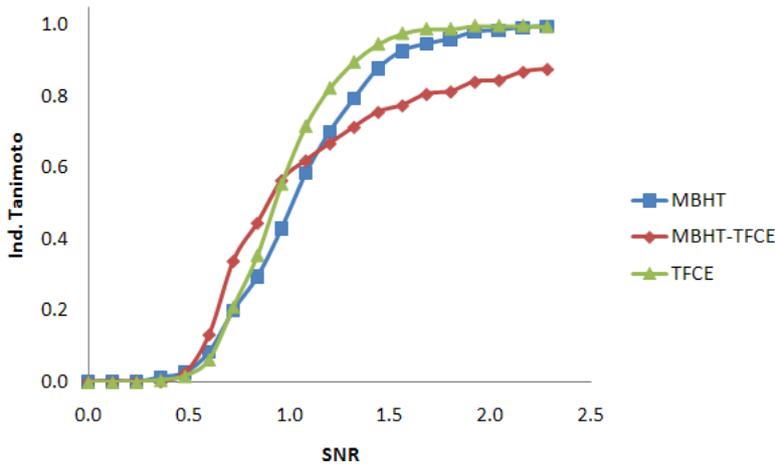


Fig. 5.1.21: Curvas de Tanimoto de: Morphology-Based Hypothesis Testing Adaptable (MBHT), Threshold Free Cluster Enhancement (TFCE) y la combinación de ambos (MBHT-TFCE) para ruido correlacionado.

En todos los métodos que involucran el histograma local ponderado fue necesario discretizar el rango de intensidad $[-4, 5]$ en $B = 100$ intervalos. Para los métodos de mediana local y moda local de la figura 5.1.23 usamos una familia \mathcal{W} de $N = 9$ Gaussianas de $\sigma = 0.0, 0.25, \dots, 2.0$; para la mediana local se usó $\sigma_I = 0.25$ y $\sigma_I = 0.50$ para la moda local. Al igual que en el caso de ruido no correlacionado la moda local muestra el mejor desempeño.

Para la media local basada en filtro bilateral, que se muestra en la figura 5.1.24 se usó una familia de $N = 13$ Gaussianas con $\sigma = 0.0, 0.25, \dots, 3.0$, $\sigma_I = 0.0$ y $\sigma_R = 1$.

En la figura 5.1.25 calculamos \bar{h}_{x, W_j} para cada $x \in L$ para hallar la mediana local. Se usó una familia de $N = 9$ Gaussianas de $\sigma = 0.0, 0.25, \dots, 2.0$, un kernel Gaussiano \mathcal{K} de $\sigma_I = 0.0$ y

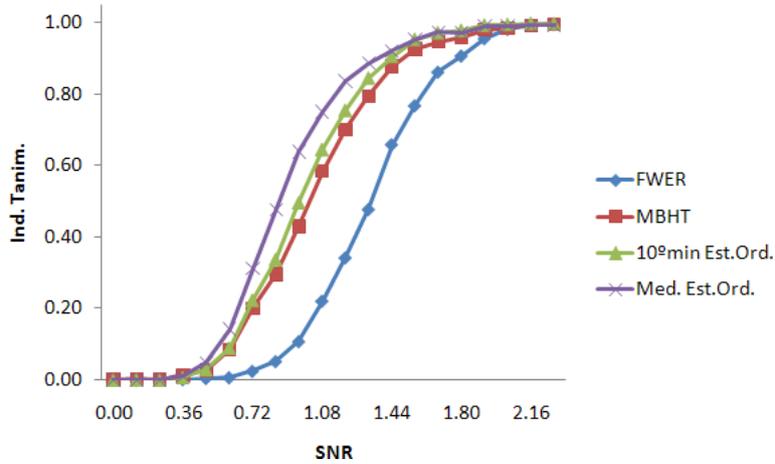


Fig. 5.1.22: Curvas índice de Tanimoto de: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT), el estadístico de orden 10 y la mediana con estadísticos de orden para ruido correlacionado.

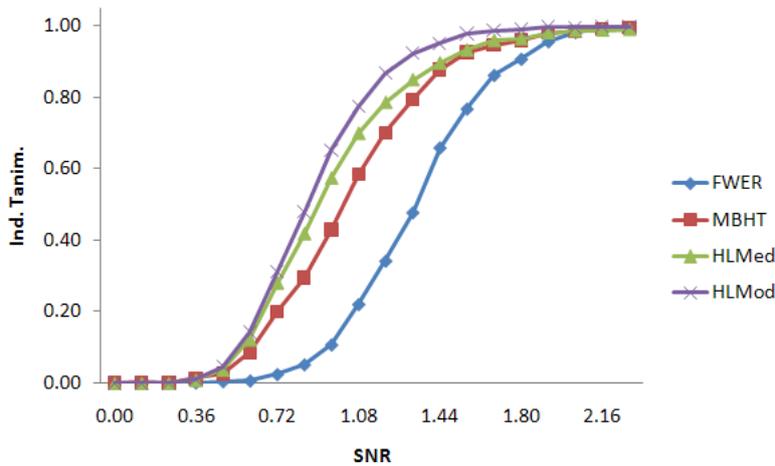


Fig. 5.1.23: Curvas del índice de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT), Mediana Local, y Moda Local para ruido correlacionado.

$\sigma_R = 1.0$ para la función \mathcal{K}_R .

La moda local en la figura 5.1.26 se obtuvo con \bar{h}_{x,W_j} para cada $x \in L$. El método se probó con distintas familias de Gaussianas \mathcal{W} y diferentes valores de σ_I . El mejor resultado que se consiguió fue con la familia de 9 Gaussianas de $\sigma = 0.0, 0.25, \dots, 2.0$, $\sigma_I = 0.25$ para el suavizado del histograma y $\sigma_R = 1$ para la función \mathcal{K}_R .

En las gráficas 5.1.24, 5.1.25 y 5.1.26 observamos que los métodos basados en filtro bilateral tienen un mejor comportamiento sustancialmente distinto a lo observado en el caso de las pruebas correspondientes de ruido no correlacionado. Todos los métodos se desempeñan mejor que el FWER y MBHT, particularmente el método de la media local.

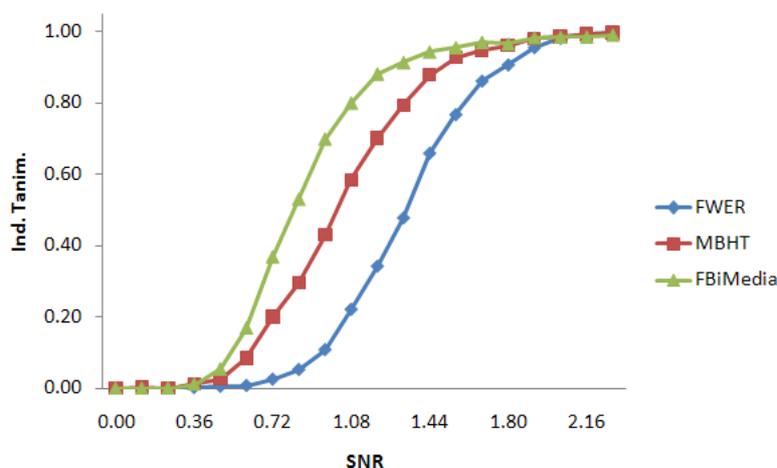


Fig. 5.1.24: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Media Local con ecuación (4.2.27) (variante de filtro bilateral) para ruido correlacionado.

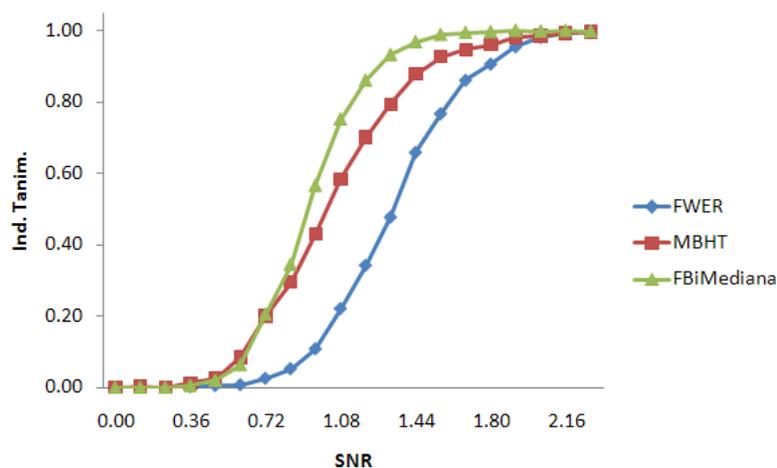


Fig. 5.1.25: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Mediana Local con expresión (4.2.30) (variante de Filtro Bilateral) para ruido correlacionado.

En la figura 5.1.27 se muestran los métodos con el mejor desempeño para el caso de ruido correlacionado. Vemos que los métodos de Mediana con Estadísticos de Orden y la Moda Local tienen un comportamiento muy similar. Notar que los cuatro métodos son comparables sin embargo el método de Media Local (variante de Filtro Bilateral) es significativamente superior al TFCE.

En la figura 5.1.28 se muestran los umbrales $\eta_{W_j}^*$ de intensidad contra el número de elementos contenidos en el kernel w_j , para cada uno de los métodos que utilizan $\tilde{h}_{x,W}(i)$. Se usó una familia de $N = 9$ Gaussianas de desviación estándar $\sigma = 0.0, 0.25, \dots, 2.0$. Se puede ver que en este

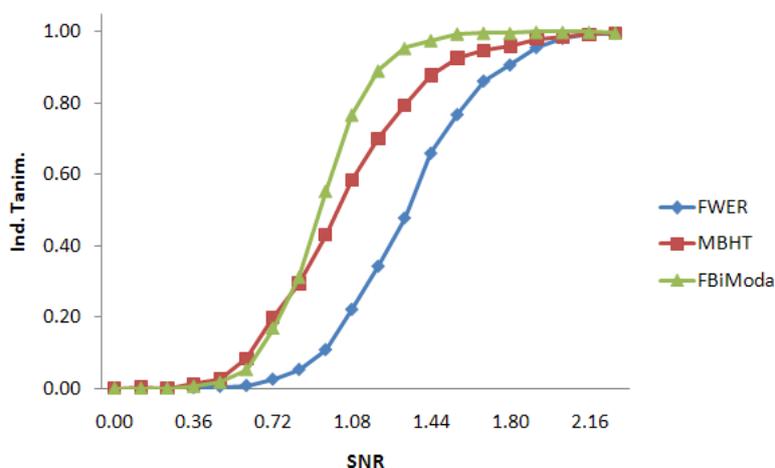


Fig. 5.1.26: Curvas de Tanimoto obtenidas con: Family Wise Error Rate (FWER), Morphology-Based Hypothesis Testing Adaptable (MBHT) y la Moda Local con expresión (4.2.34) (variante de filtro bilateral) para ruido correlacionado.

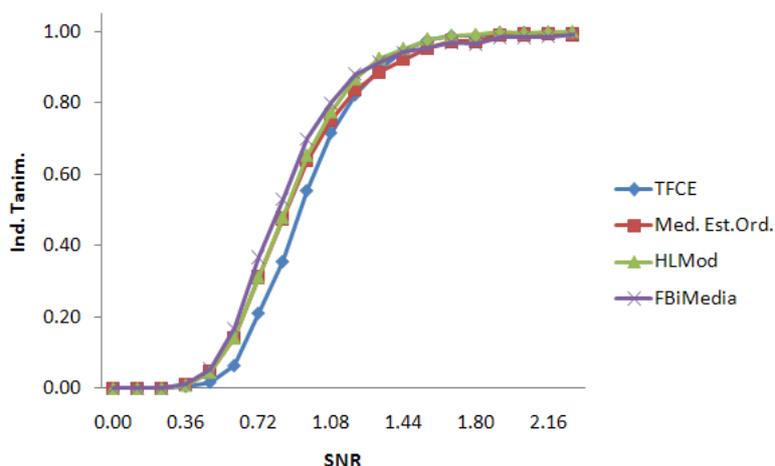


Fig. 5.1.27: Curvas de Tanimoto obtenidas con Threshold Free Cluster Enhancement (TFCE), Mediana con estadístico de orden (MedEO), Moda Local con histograma local ponderado (HLMMod) y variante de Filtro Bilateral (FBiMedia) para ruido correlacionado.

caso las gráficas de los métodos basados en filtro bilateral decrecen de manera semejante a la mediana y moda local, a excepción de la moda local basada en filtro bilateral.

Las figuras 5.1.29 a 5.1.37 exhiben las regiones de activación estimadas con los métodos MTSTCS, FWER, TFCE, MBHT, MBHT-TFCE, estadísticos de orden $n = 10$, Mediana local con estadísticos de orden, Moda Local (con $\tilde{h}_{x,W}$) y Media local con $\bar{h}_{x,W}$ (variante de filtro bilateral).

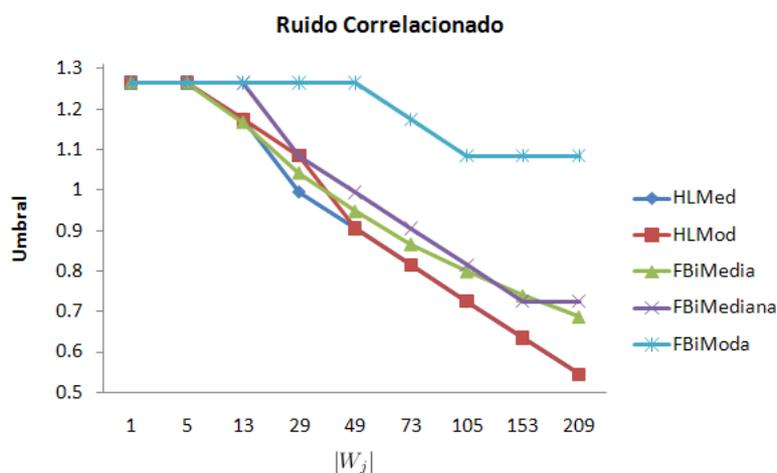


Fig. 5.1.28: Gráfica del valor del umbral $\eta_{W_j}^*$ en relación con el número de elementos que contiene el kernel w_j para cada uno de los métodos basados en el histograma local ponderado.

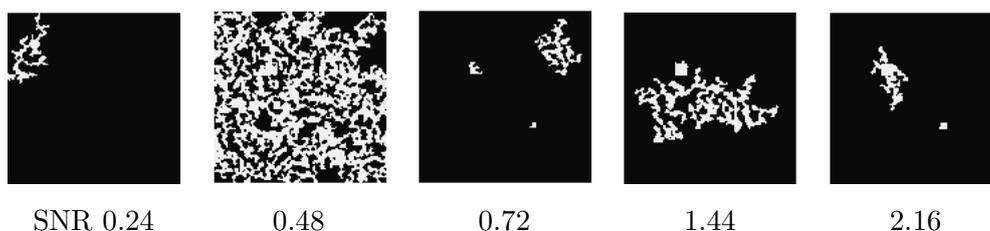


Fig. 5.1.29: Región de activación estimada con el método MTSTCS para diferentes valores de SNR, para ruido correlacionado.

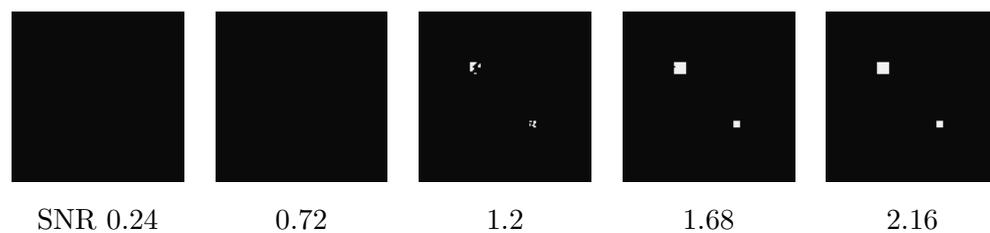


Fig. 5.1.30: Estimación de la región de activación con el método FWER para diferentes valores de SNR, para ruido correlacionado.

5.2. Datos Sintéticos fMRI

Este experimento considera el problema de detectar RA en conjuntos de volúmenes fMRI obtenidos en experimentos específicos. En este contexto, es importante usar un conjunto de datos que sea realista. Esto es, que las correlaciones espaciales y temporales sean similares a las que se presentan en la configuración experimental actual. A la vez es necesario conocer la localización de la región de activación. De esta manera nos es posible realizar una evaluación

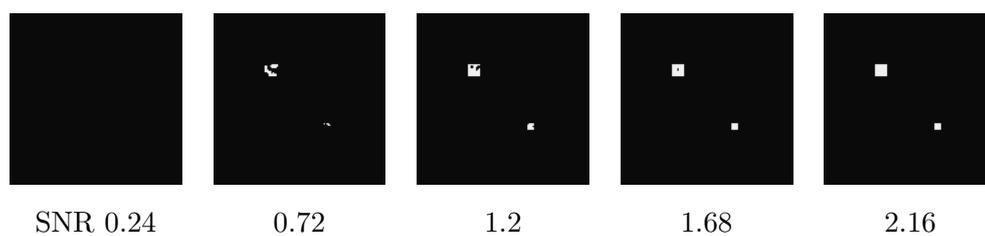


Fig. 5.1.31: Estimación de la región de activación con el método TFCE para diferentes valores de SNR, para ruido correlacionado.

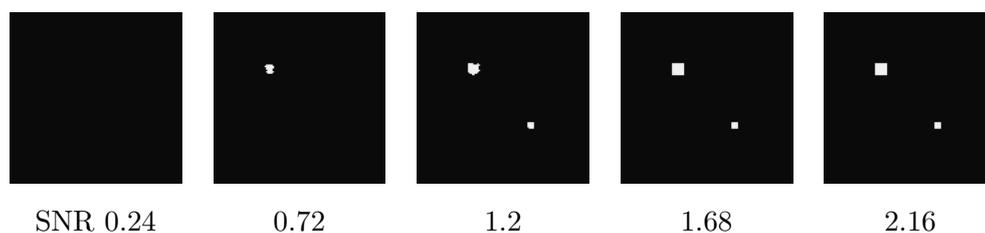


Fig. 5.1.32: Región de activación estimada con el método MBHT para diferentes valores de SNR, para ruido correlacionado.

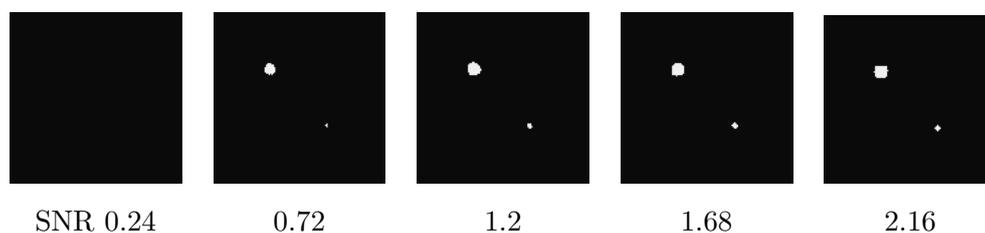


Fig. 5.1.33: Región de activación estimada con el método MBHT-TFCE para diferentes valores de SNR, para ruido correlacionado.

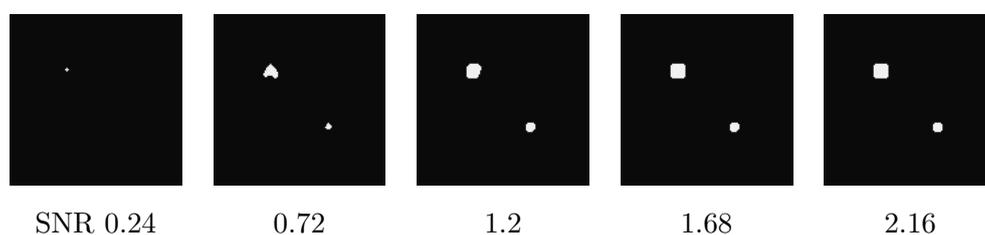


Fig. 5.1.34: Región de activación estimada con el 10^2 mín para diferentes valores de SNR.

precisa del método. Para ello, se construyó un diseño aleatorio de simulación, en el cual se obtienen 42 escaneos durante la presentación de un supuesto estímulo (“on”scans). Cada escaneo es asignado aleatoriamente de un total de 84 escaneos. De acuerdo con el modelo estándar de convolución ([9, Friston et al., (1994a)]), la intensidad registrada en el voxel x del escaneo k se

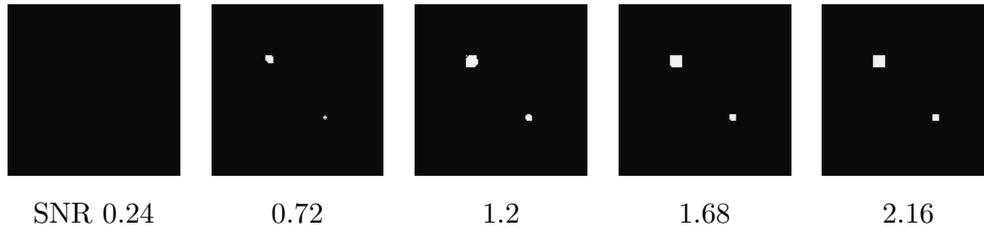


Fig. 5.1.35: Región de activación estimada con Mediana Local de Estadísticos de Orden para diferentes valores de SNR.

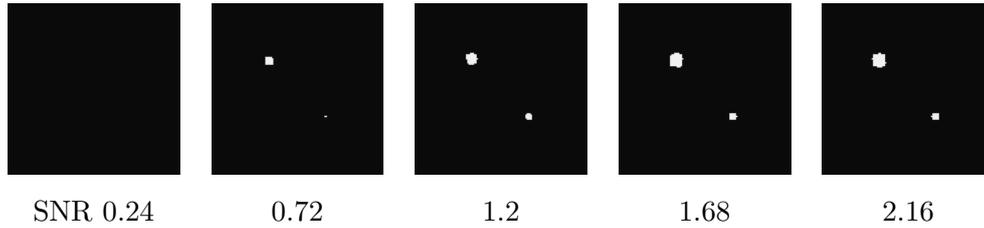


Fig. 5.1.36: Resultados obtenidos con la moda local para diversos valores de SNR.

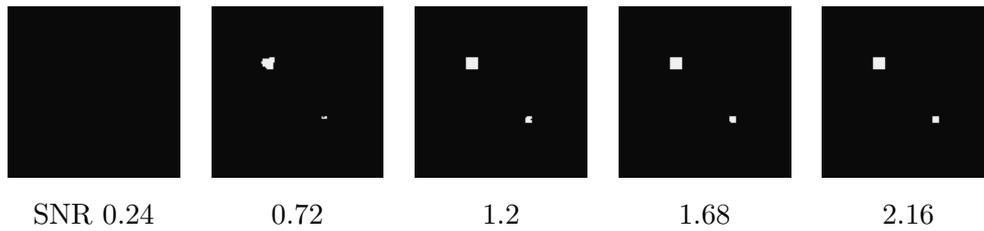


Fig. 5.1.37: Resultados obtenidos con la Media Local (variante de Filtro Bilateral) para diversos valores de SNR.

modeló con:

$$Y_k(x) = \gamma_0(x) + \gamma_1(x) \sum_j z_j h(k - j) + n_k(x) \quad (5.2.1)$$

donde z es el vector de estímulos (esto es, $z_j = 1$ si un estímulo se presentó en el tiempo j , y $z_j = 0$ en otro caso), h es la Función de Respuesta Hemodinámica (HRF, Hemodynamic Response Function), n es un proceso de ruido, y $\gamma_0(x)$, $\gamma_1(x)$ son coeficientes que dependen de la presencia de una activación en el voxel x .

Para generar un proceso de ruido que sea más apegado a la realidad, usamos el conjunto de datos fMRI Starplus (mismo que usaron en [16, Marroquín et al., (2011)]), el cual originalmente fue recolectado por M. Just in Carnegie Mellon University's Center for Cognitive Brain Imaging (CCBI) ([17, Keller et al., (2001)];[20, Mitchell et al., (2004)]). De este conjunto, solamente se consideraron las imágenes fMRI que fueron tomadas bajo una condición de referencia (base), en la cual el sujeto simplemente estaba en reposo. Los datos fueron preprocesados para eliminar

los artefactos de movimiento debido a movimiento de la cabeza, desviación de la señal, y otras causas, usando el programa FIASCO ([8, Eddy, (1998)]). Todos los valores de activación de los píxeles se midieron sobre regiones definidas anatómicamente de interés (ROIs) dentro del cerebro de cada sujeto. En los experimentos que aquí se presentan consideramos un conjunto S_0 de 512 imágenes y se generaron MC corridas seleccionando al azar, para cada corrida, 84 volúmenes (con reemplazo) para el conjunto S_0 . Solamente se consideró un corte por volumen para este experimento, consistente de 64×64 voxeles; esto se hizo por razones computacionales y también porque solamente están disponibles 8 cortes por volumen, lo que significa que prácticamente no existe correlación espacial entre cortes, además los elementos estructurales (esferas o Gaussianas) que se usan podrían contener solamente un corte.

El coeficiente γ_0 se estableció como cero, mientras que γ_1 se igualó a un nivel de amplitud fijo a para los voxeles dentro de un cuadrado de 5×5 localizado en la corteza occipital derecha (ver la 5.2.1). La HRF h se generó usando la función incluida en el software SPM ([11, Friston et al., (2007)]).



Fig. 5.2.1: La imagen de la izquierda muestra la región de activación usada en las pruebas para datos sintéticos fMRI. La imagen de la derecha es el campo muestral obtenido con un nivel de activación igual a 9.

Para realizar la detección se considerara que la respuesta hemodinámica sigue la HRF canónica, y por lo tanto la hipótesis nula de no presentar activación en el voxel x es equivalente a $H_{0x} : \gamma_1(x) = 0$. En este caso, el campo t está formado simplemente por el conjunto de estadísticos locales F (aunque no se supone que sigan una distribución F).

Las distribuciones empíricas F_{θ_w} y F_g se obtuvieron generando permutaciones del vector de estímulos (esto es, de las etiquetas de escaneos). Tomar en cuenta que no se permutan los volúmenes a sí mismos, sino solamente las etiquetas de estímulos, la prueba de permutación sigue siendo válida independientemente de la estructura de la correlación temporal de los datos, ya que las etiquetas son intercambiables bajo H_0 ([27, Raz et al., (2003)]). Ya que el total del número de permutaciones ($C_{42}^{84} = \frac{84!}{2(42!)}$) es muy grande, se usó un número fijo de 2000 permutaciones aleatorias. Solo se probaron los métodos: 10^o mín Est. Ord., Mediana con Est. de Ord., Moda con histograma local ponderado, y Media (variante de Filtro Bilateral). Estos

métodos mostraron un mejor desempeño en el experimento de simulación de campos Gaussianos. Se incluyen los métodos FWER y MBHT para realizar una comparación entre ellos. Los métodos se corrieron con los mismos parámetros que se usaron para ruido correlacionado en el caso de campos aleatorios Gaussianos.

Índice Tanimoto Promedio					
FWER	MBHT	Med. Est.Ord.	10 ^e mín Est.Ord.	HLMod	FBiMedia
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.051519	0.232000	0.246008	0.228337	0.191060	0.214889
0.488237	0.805747	0.871041	0.871707	0.858571	0.794965
0.969362	0.944494	0.960988	0.972872	0.993109	0.927012
1.000000	0.979723	0.975643	0.984793	0.992505	0.940450
1.000000	0.991874	0.996667	1.000000	0.994945	0.958506

Tabla 5.2.1: Índice de Tanimoto obtenido en los experimentos de fMRI.

La tabla 5.2.1 muestra el índice de Tanimoto promedio que se obtuvo para cada uno de los métodos en los experimentos de fMRI. En las figuras 5.2.3 a 5.2.6 se muestran las gráficas de estos valores.

Probabilidad de Deteccion					
FWER	MBHT	Med. Est.Ord.	10 ^e mín Est.Ord.	HLMod	FBiMedia
0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
0.360000	0.440000	0.520000	0.500000	0.420000	0.500000
0.960000	0.980000	1.000000	1.000000	1.000000	0.980000
1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Tabla 5.2.2: Probabilidad de Detección en los experimentos de fMRI.

La tabla 5.2.2 muestra la probabilidad de detección de cada método para los datos sintéticos de fMRI. En las figuras 5.2.7 a 5.2.10 pueden verse las gráficas de estos resultados.

La figura 5.2.2 muestra las regiones de activación estimadas, que se hallaron con cada método en el experimento con datos sintéticos de fMRI.

En las figuras 5.2.3 a 5.2.6 no se observan cambios significativos en el comportamiento de los métodos.

En las figuras 5.2.7 a 5.2.10 tampoco se observa una diferencia significativa en la probabilidad de detección.

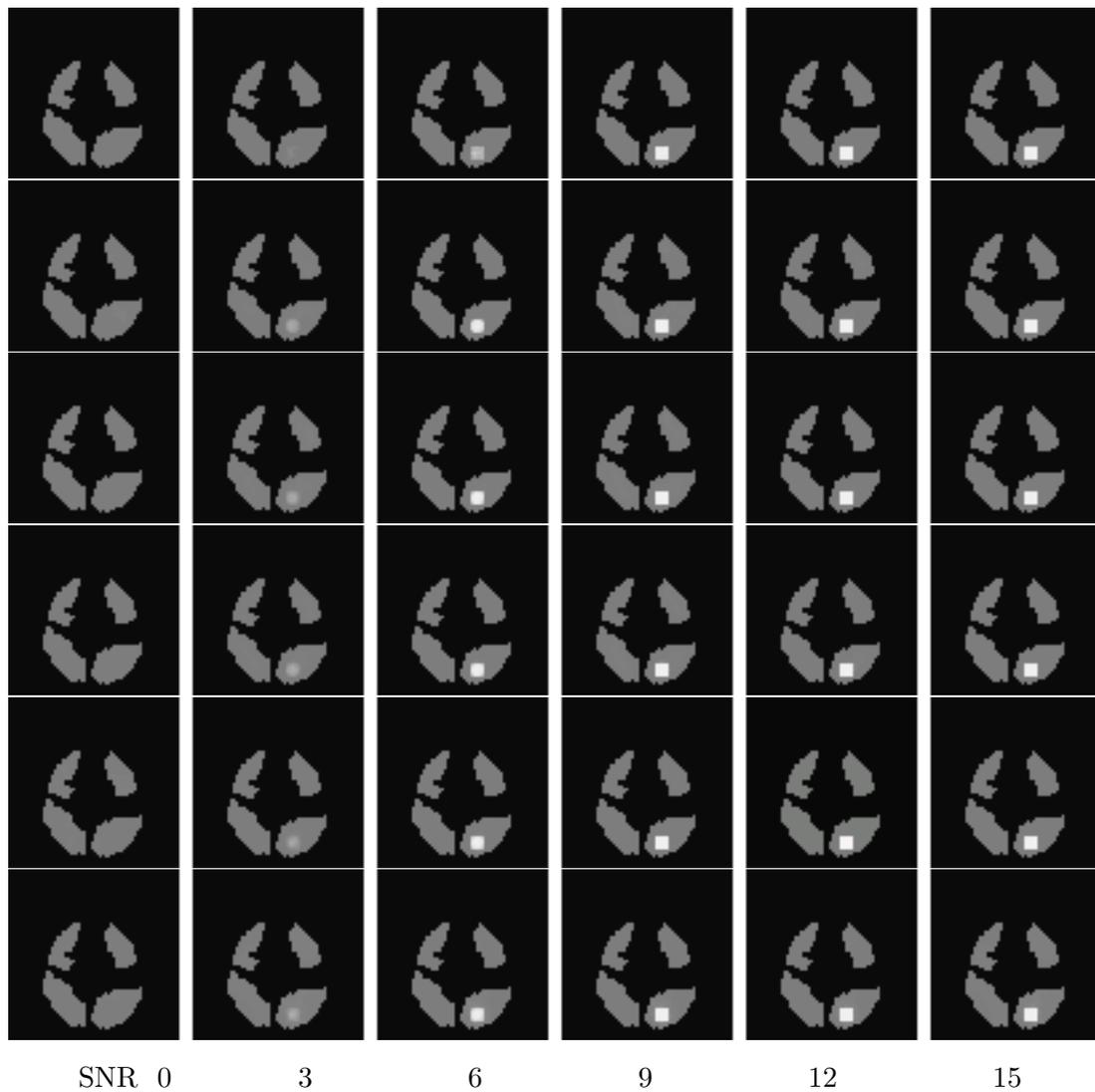


Fig. 5.2.2: Los resultados se muestran por renglón en el siguiente orden: En el primer lugar el FWER seguido por el MBHT; en el número tres está el estadístico de orden $n = 10$; en la posición cuatro, la mediana local con estadísticos de orden; en penúltimo lugar aparece la Moda Local (histograma local ponderado) y por último la Media Local (variante de Filtro Bilateral).

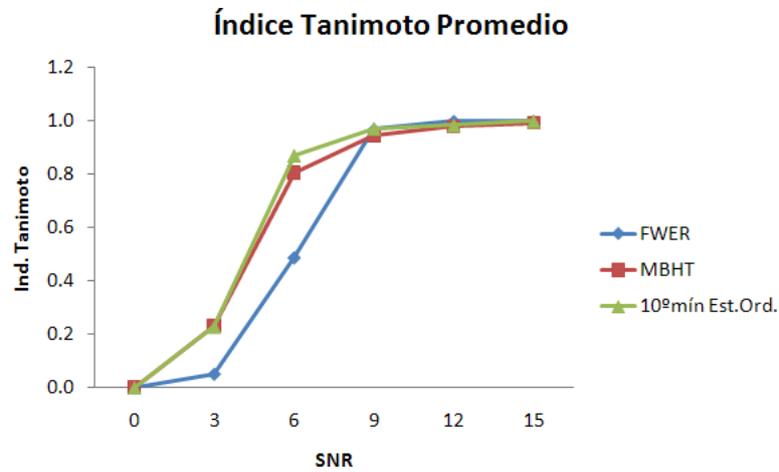


Fig. 5.2.3: Índice de Tanimoto obtenido en los experimentos de fMRI con los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Estadístico de Orden $n = 10$.

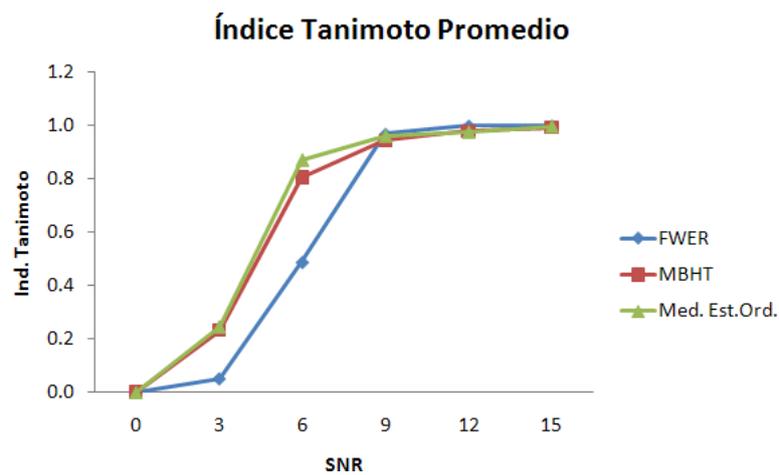


Fig. 5.2.4: Índice de Tanimoto obtenido en los experimentos de fMRI con los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Mediana con Estadístico de Orden.

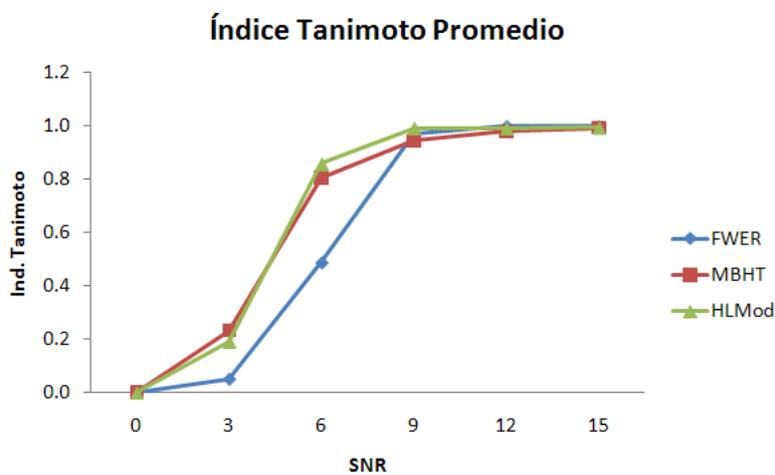


Fig. 5.2.5: Índice de Tanimoto obtenido en los experimentos de fMRI con los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Moda con histograma local ponderado suavizado.

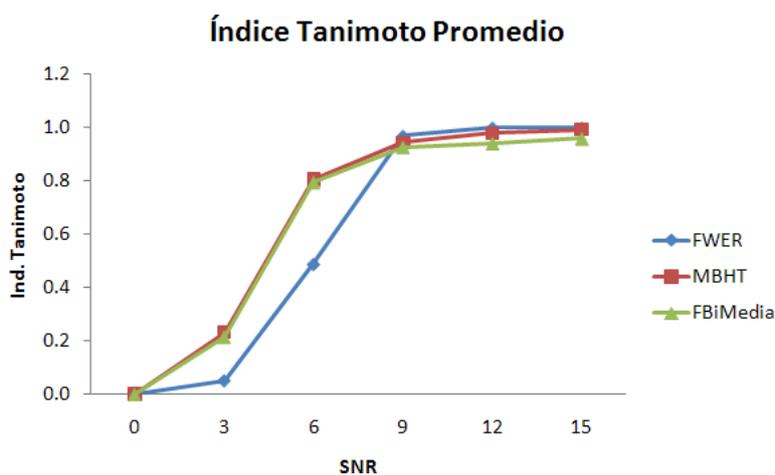


Fig. 5.2.6: Índice de Tanimoto obtenido en los experimentos de fMRI con los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Media (variante de Filtro Bilateral).

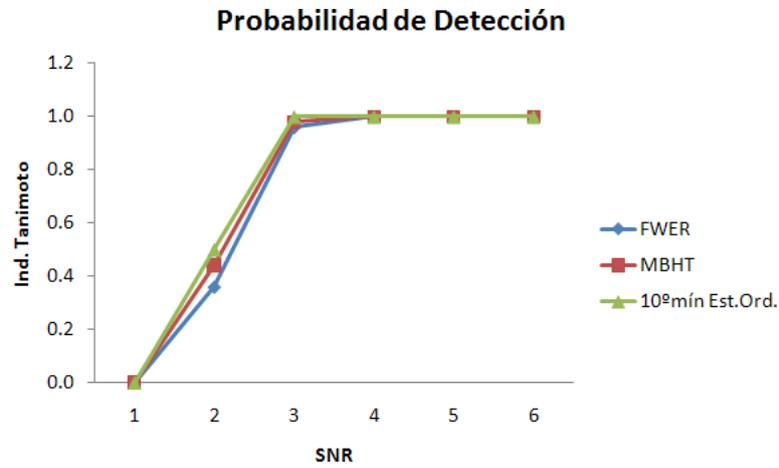


Fig. 5.2.7: Probabilidad de detección de los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Estadístico de Orden $n = 10$.

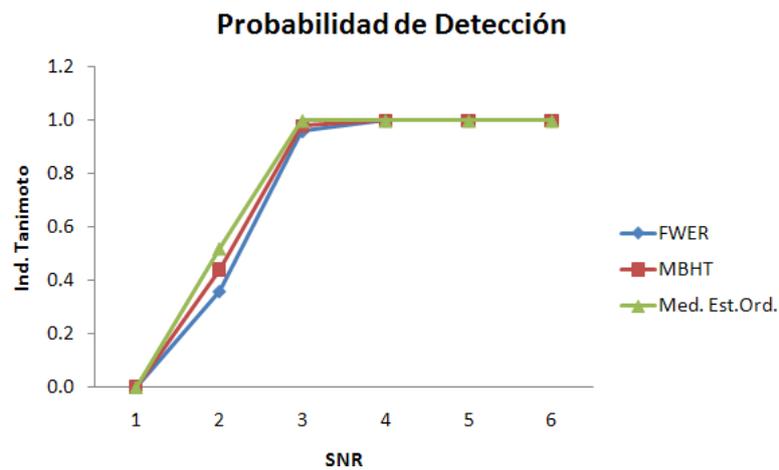


Fig. 5.2.8: Probabilidad de detección de los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Mediana con Estadístico de Orden.

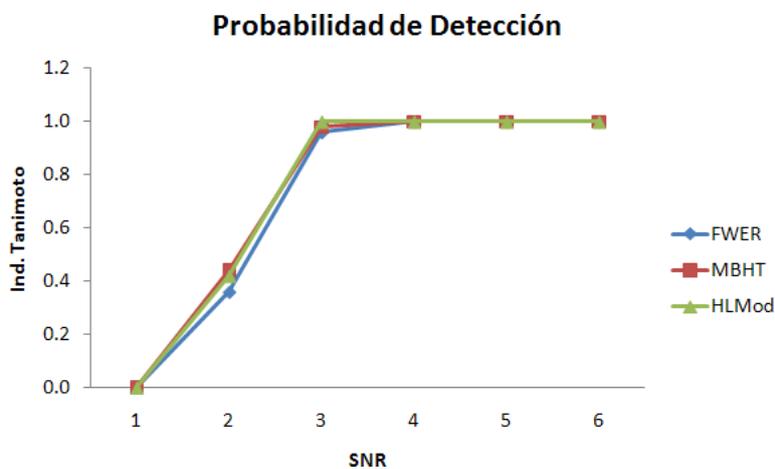


Fig. 5.2.9: Probabilidad de detección de los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Moda con histograma local ponderado suavizado.

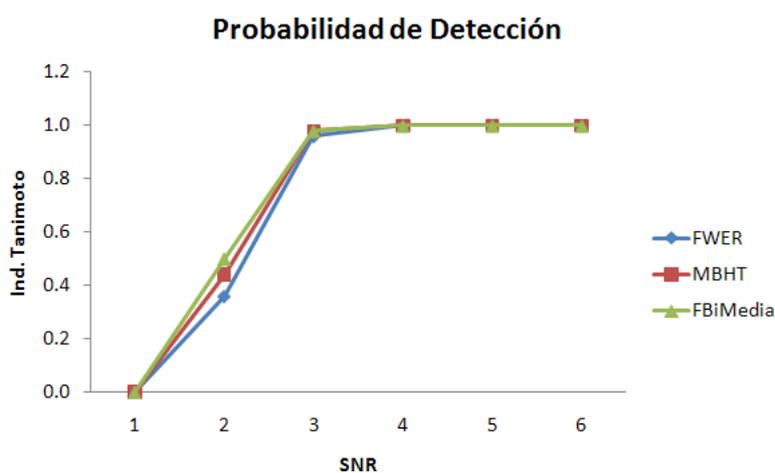


Fig. 5.2.10: Probabilidad de detección de los métodos Family-Wise Error Rate (FWER), Morfology Based Hypothesis Testing (MBHT) y Media (variante de Filtro Bilateral).

Conclusiones

En este trabajo de tesis se presentaron 6 nuevos procedimientos estadísticos para la detección de activaciones en campos aleatorios discretos: método basado en estadísticos de orden, métodos basados en histograma local ponderado suavizado (mediana y moda local), métodos basados en filtro bilateral (media, mediana y moda). Dichos procedimientos están relacionados al método MBHT propuesto por [16, Marroquin et. al (2011)]. Mediante diversos ejemplos se mostró que pueden ser utilizados para enfrentar el problema de comparaciones múltiples en neuroimagen.

En [16, Marroquin et. al (2011)] se muestra que el algoritmo MBHT ofrece resultados superiores a los métodos estándar como el FWER. Más aún, se muestra que el algoritmo MBHT mejora la detección de regiones de activación y la estimación de su extensión, además de poseer propiedades deseables de control de errores. No obstante, este algoritmo presenta dificultades para detectar activaciones de forma adecuada cuándo el nivel y la extensión de la activación son pequeños. Una forma de atacar estos problemas es reducir la magnitud de los umbrales de rechazo.

Los algoritmos que se desarrollaron aquí tenían como objetivo principal atacar las deficiencias del algoritmo MBHT. Para ello, nuestro enfoque incorpora estadísticos de orden y medidas de tendencia central locales (basados en un elemento estructural centrado en el voxel \mathbf{x} , o en los elementos pertenecientes a un kernel Gaussiano centrado en \mathbf{x} , respectivamente).

En el trabajo se presentaron varios ejemplos de las aplicaciones de los procedimientos desarrollados sobre datos sintéticos. En el caso de simulación de campos Gaussianos, algunos de nuestros algoritmos demostraron tener un mejor desempeño que el método MBHT. La mejoría fue evidente en los ejemplos en los que los niveles de las activaciones eran pequeños. El método de la moda local tuvo el mejor desempeño general: el mejor desempeño absoluto en el caso de ruido no-correlacionado y comparable con el mejor para ruido correlacionado (media local del filtro bilateral).

En el caso de datos sintéticos fMRI, nuestros algoritmos generan resultados comparables con los obtenidos por el método MBHT. Sin embargo, ninguno de ellos fue capaz de obtener resul-

tados significativamente mejores.

Una de las ventajas importantes de los métodos propuestos es que son muy generales, al igual que el MBHT. Esto se debe principalmente a que no hacen supuestos fuertes sobre la distribución de los campos aleatorios. Es cierto que nuestros métodos requieren un poco más de información para su implementación, por ejemplo, el rango de valores discretizado del campo aleatorio; no obstante, esto no reduce su potencial.

Los métodos basados en histograma local ponderado suavizado requieren de un mayor costo computacional. Sin duda lo más costoso es llevar a cabo la primera etapa de los algoritmos; sin embargo, ésta sólo se realiza una vez para hallar los umbrales $\eta_{W_j}^*$ (con los parámetros adecuados). Por ejemplo, para calcular los umbrales $\eta_{W_j}^*$, el método de Estadísticos de Orden $n = 10$ tarda aproximadamente 15 minutos, mientras que la Moda Local (con histograma local ponderado suavizado) tarda alrededor de 160 minutos.

En general se observó que los métodos que emplean elementos estructurales tienden a sobrestimar la RA un poco más que aquellos que utilizan familias de Gaussianas. Esto se debe, a que los últimos asignan pesos a la información que aportan los voxeles vecinos dependiendo de su distancia al voxel central.

Los métodos nuevos que se proponen aquí, utilizan la distribución del máximo y una escala de significancia para mantener bajo control la tasa de falsos positivos. Dado que están basados en operadores de ventana con soporte local, tienen control de error de localización acotado por el radio del soporte del operador (por ejemplo, 4σ en el caso de la moda local). De hecho, en el caso de los métodos basados en histograma local ponderado suavizado, esta cota está sobrestimada. Esto se debe a que los voxeles vecinos y , reciben un peso dependiendo de su distancia al voxel x , cuanto más lejos se encuentran de x menor es su peso. Entonces, los voxeles “cercaños” a la frontera del soporte del kernel Gaussiano W_x que caen dentro de la región de activación (considerados como outliers), por lo general no afectarán el comportamiento del operador de ventana. Es por ello que el error de localización está acotado por un $\epsilon < 4\sigma$. La demostración rigurosa de este hecho, así como hallar una mejor cota para el error de localización, van más allá de los propósitos de este trabajo.

Si se desea tener además control de error basado en cluster, habría que definir, como estimador de la región de activación, la dilatación morfológica de la región obtenida por los métodos del capítulo 4, usando como elemento estructurante el soporte del operador. Esto, sin embargo, degradaría el desempeño de estos métodos, y aumentaría la cota del error de localización de los mismos, por lo que no es recomendable hacerlo.

Como trabajo futuro se espera encontrar hallar mejores cotas para el error de localización. Realizar experimentos realistas donde los clusters de activación difieran significativamente de los clusters formados por ruido. Así como utilizar otras mediciones para evaluar el desempeño de los métodos, que nos ayuden a distinguir entre la tasa de falsos positivos y falsos negativos que

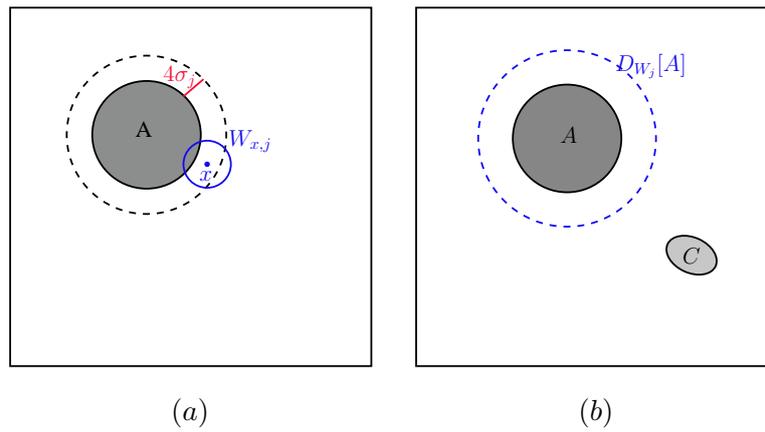


Fig. 6.0.1: (a) Si el voxel x es tal que $d(x, A) \leq 4\sigma_j$, entonces el soporte del kernel Gaussiano $W_{x,j}$ se interseca con A ; (b) La $P(C \text{ activo} \mid |C \cap D_{W_j}[A] = \emptyset) < \alpha$, pero se incrementa la tasa de falsos positivos, pues se sobrestima la región de activación A .

se presentan en la práctica.

Bibliografía

- [1] Abdi, H., 2007. *Bonferroni and Sidák corrections for multiple comparisons*. In: Salkind, N.J. (Ed.), *Encyclopedia of Measurement and Statistics*. Sage, Thousand Oaks, CA. 1.2
- [2] Benjamini, Y., Heller, R., 2007. *False discovery rates for spatial signals*. *J. Am. Stat. Assoc.* 102 (480), 1272–1281. 1.2
- [3] Benjamini, Y., Hochberg, Y., 1995. *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. *J. R. Stat. Soc. B* 57, 289–300. 1.2
- [4] Bullmore, E.T., Suckling, J., Overmeyer, S.R., Taylor, E., Brammer, M.J., 1999. *Global, voxel and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain*. *IEEE Trans. Med. Imaging* 18 (1), 32–42. 1.2, 2.3, 3.4
- [5] George Casella & Roger L. Berger, 2002. *Statistical Inference*. 2nd ed., Duxbury Thomson Learning. 1.1
- [6] Chumbley, J.R., Friston, K.J., 2009. *False discovery rate revisited: FDR and topological inference using Gaussian random fields*. *Neuroimage* 44 (1), 62–70. 1.2
- [7] Chumbley, J.R., Worsley, K., Flandin, G., Friston, K., 2009. *Topological FDR for neuroimaging*. *Neuroimage* 49, 3057–3064. 1.2
- [8] Eddy, W., 1998. *The challenge of functional magnetic resonance imaging*. *J. Comput. Gr. Stat.* 8, 545–558. 5.2
- [9] Friston, K.J., Jezzard, P., Turner, R., 1994a. *Analysis of functional MRI time series*. *Hum. Brain Mapp.* 1 (2), 153–171. 5.2
- [10] Friston, K.J., Worsley, K.J., Frackowiak, R.S.J., Mazziotta, J.C., Evans, A.C., 1994b. *Assessing the significance of focal activations using their spatial extent*. *Hum. Brain Mapp.* 1, 214–220. 1.2, 2.2

-
- [11] Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D. (Eds.), 2007. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press. 5.2
- [12] Genovese, C.R., Lazar, N., Nichols, T.E., 2002. *Thresholding of statistical maps in functional neuroimaging using the false discovery rate*. *Neuroimage* 15 (4), 870–878. 1.2
- [13] Hayasaka, S., Nichols, T.E., 2004. *Combining voxel intensity and cluster extent with permutation test framework*. *Neuroimage* 23, 54–63. 1.2
- [14] Heller, R., Stanley, D., Yekutieli, D., Rubin, N., Benjamini, Y., 2006. *Cluster-based analysis of fMRI data*. *Neuroimage* 33, 599–608. 1.2
- [15] Holmes, A.P., Blair, R.C., Watson, J.D.G., Ford, I., 1996. *Nonparametric analysis of statistic images from functional mapping experiments*. *J. Cereb. Blood Flow Metab.* 16, 722. 1.3
- [16] José L. Marroquín, Rolando J. Biscay, Salvador Ruíz-Correa, Alfonso Alba, Roxana Ramírez, Jorge L. Armony, 2011. *Morphology-based hypothesis testing in discrete random fields: A non-parametric method to address the multiple-comparison problem in neuroimaging*. *NeuroImage* 56 (2011) 1954–1967. 3, 3.1.1, 3.1.1, 3.1.1, 3.1.1, 3.1.2, 3.1.2, 3, 3.3, 3.5, 5.2, 6
- [17] Keller, T.A., Just, M.A., Stenger, V.A., 2001. *Reading Span and the Time-Course of Cortical Activation in Sentence–Picture Verification*. Annual Convention of the Psycho-nomic Society, Orlando, FL. 5.2
- [18] Langer, D.R.M., Jansen, J.F.A., Backes, W.H., 2007. *Enhanced signal detection in neuroimaging by means of regional control of the global false discovery rate*. *Neuroimage* 38, 43–56. 1.2
- [19] Lehmann, E.L. and Romano, J.P., 2005. *Testing statistical hypothesis*. Third Edition. Springer Science+Business Media, Inc., New York. 1.3, 2, 5
- [20] Mitchell, T.M., Hutchinson, R., Niculescu, R.S., Pereira, F., Wang, X., Just, M., Newman, S., 2004. *Learning to decode cognitive states from brain images*. *Mach. Learn.* 57 (1–2), 145–175. 5.2
- [21] Nichols, T.E., 2002. *FWE-Corrected Inference: Parametric Conservativeness and Nonpara-metric Alternatives*. Proceedings of the Joint Statistical Meetings, New York. 1.2
-

-
- [22] Nichols, T.E., Hayasaka, S., 2003. *Controlling the familywise error rate in functional neuroimaging: a comparative review*. Stat. Methods Med. Res. 12, 419–446. 1.2
- [23] Pesarin, F., 2001. *Multivariate Permutation Tests with Applications in Biostatistics*. John Wiley and Sons, New York. 3.1.2
- [24] Poline JB, Mazoyer BM, 1993. *Analysis of individual positron emission tomography activation maps by detection of high signal-to-noise-ratio pixel clusters*. J Cereb Blood Flow Metab 13:425437. 2.2
- [25] Poline, J.B., Worsley, K.J., Evans, A.C., Friston, K.J., 1997. *Combining spatial extent and peak intensity to test for activations in functional imaging*. Neuroimage 5, 83–96. 1.2
- [26] Purdon, P., Weisskoff, R., 1998. *Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI*. Hum. Brain Mapp. 6, 239249. 1.2
- [27] Raz, J., Zheng, H., Ombao, H., Turetsky, B., 2003. *Statistical tests based on experimental randomization*. Neuroimage 19, 226–232. 2, 5.2
- [28] Roland P.E., Levin B., Kawashima R., Ackerman S., 1993. *Three dimensional analysis of clustered voxels in 150-butanol brain activation images*. Hum Brain Mapping 1:3-19. 2.2
- [29] Tomasi, C., and Manduchi, R., 1998. *Bilateral filtering for gray and color images*. Proceedings of the IEEE International Conference on Computer Vision, 839-846. 4.2.3
- [30] Sergerie, K., Chochol, C., Armony, J.L., 2008. *The role of amigdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies*. Neurosci. Biobehav. Rev. 32, 811–830. 1.2
- [31] Serra, J., 1984. *Image Analysis and Mathematical Morphology*. Ac. Press. 3, 3.1.1
- [32] Smith, S.M., Nichols, T.E., 2009. *Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference*. Neuroimage 44, 83–98. 1.2, 2.4, 2.4, 3.5
- [33] Soille, P., 2004. *Morphological Image Analysis*. Springer. 3
- [34] Worsley, K., Evans, A., Marrett, S., Neelin, P., 1992. *A three-dimensional statistical analysis for CBF activation studies in human brain*. J. Cereb. Blood Flow Metab. 12, 900918. 1.2
-

- [35] Zarahn, E., Aguirre, G., D'Esposito, M., 1997. *Empirical analyses of BOLD fMRI statistics*. Neuroimage 5, 179-197. [1.2](#)
- [36] Zhang, H., Nichols, T.E., Johnson, T.D., 2009. *Cluster mass inference via random field theory*. Neuroimage 44, 51-61. [1.2](#)
-