Regularized Hypothesis Testing in Random Fields with Applications to Neuroimaging

Pruebas de Hipótesis Regularizadas en Campos Aleatorios con Aplicaciones a Neuroimágenes

Oscar S. Dalmau-Cedeño, Dora E. Alvarado-Carrillo, José Luis Marroquín

Centro de Investigación en Matemáticas, CIMAT A.C.

ABSTRACT

The task of determining for which elements of a random field (e.g., pixels in an image) a certain null hypothesis may be rejected is a relevant problem in several scientific areas. In the current contribution, we introduce a new method for performing this task, the regularized hypothesis testing (RHT) method, focusing on its use in neuroimaging research. RHT is based on the formulation of the hypothesis testing task as a Bayesian estimation problem, with the previous application of a Markovian random field. The latter allows for the incorporation of local spatial information and considers different noise models, including spatially correlated noise. In tests on synthetic data showing regular activation levels on uncorrelated noise fields, RHT furnished a true positive rate (TPR) of 0.97, overcoming the state-of-the-art morphology-based hypothesis testing (MBHT) method and the traditional family-wise error rate (FWER) method, which afforded 0.93 and 0.58, respectively. For fields with highly correlated noise, the TPR provided by RHT was 0.65, and by MBHT and FWER was 0.35 and 0.29, respectively. For tests utilizing real functional magnetic resonance imaging (fMRI) data, RHT managed to locate the activation regions when 60% of the original signal were removed, while MBHT located only one region and FWER located none.

KEYWORDS: Regularized hypothesis test; Markovian random fields; Bayesian estimation; functional Magnetic resonance imaging.

RESUMEN

En varias áreas científicas aparece el problema de determinar los elementos de un campo aleatorio (por ejemplo, píxeles en una imagen) en los que se puede rechazar una cierta hipótesis nula. En este artículo presentamos un nuevo método para realizar esta tarea, centrado en aplicaciones para investigación de neuroimagen. Nuestra propuesta se basa en la formulación de pruebas de hipótesis como un problema de estimación Bayesiana, usando como a priori un campo aleatorio Markoviano, que permite incorporar información espacial local y considera diferentes modelos de ruido, incluido el ruido correlacionado espacialmente. Para pruebas en datos sintéticos con niveles de activación regulares sobre campos de ruido no correlacionado, nuestro método obtiene una tasa de verdaderos positivos (TPR) de 0.97, superando al método del estado del arte MBHT y al método de control FWER que obtienen 0.93 y 0.58 respectivamente; para campos con ruido altamente correlacionado, nuestro método obtiene un TPR de 0.65, mientras que MBHT y FWER obtienen 0.35 y 0.29 respectivamente. Para pruebas con datos reales de fMRI, nuestro método localiza las regiones de activación cuando removemos 60% de la señal original, mientras que MBHT no localiza región alguna y FWER localiza una de las dos regiones.

PALABRAS CLAVE: Prueba de hipótesis regularizada; campo aleatorio Markoviano; estimación Bayesiana; Imágenes de Resonancia Magnética Funcional.

Correspondencia

DESTINATARIO: **Oscar Susano Dalmau Cedeño**

INSTITUCIÓN: **Centro de Investigación en Matemáticas, CIMAT A.C.**

DIRECCIÓN: **Jalisco S/N, Col. Valenciana, C. P. 36023, Guanajuato, Guanajuato, México**

CORREO ELECTRÓNICO: **dalmau@cimat.mx**

INTRODUCTION

In areas of scientific research where imaging is involved (e.g., neuroimaging, remote sensing, etc.), it is often necessary to test statistical hypotheses at each element of a 2 or 3-dimensional field of sites (pixels or voxels). The purpose is to determine the set of sites at which the response to a given experiment may be different from baseline, or whether it is significantly correlated with another parameter.

For instance, researchers in the area of neuroscience typically conduct studies to identify the area of the brain responsible for a certain cognitive task. The experiments are composed of stimulus and rest periods applied to a single subject or several people [1] [2] [3] [4] [5], applying a functional imaging approach such as positron emission tomography (PET) or functional magnetic resonance imaging (fMRI). Subsequently, the regions of voxels with a significant degree of activation have to be detected by performing simultaneous hypothesis tests over 2 or 3-dimensional measurements.

Because hundreds of thousands of comparisons are made at the same time, the well-known problem of multiple comparisons appears [6] [7]. The researcher is thus obliged to seek a solution to the resulting increase in the percentage of false positives (type I errors). A popular family of approaches to deal with this problem are the so-called pointwise (PW) methods, which utilize some type of thresholding technique to control the family-wise error rate (FWER) [6] [8]. Although they present a simple and easy to interpret solution, there is a high rate of type I errors.

Some authors address the problem through the Gaussian random fields (GRF) theory [9] [10], under the assumption that the spatial correlation of the data is known or can be estimated. Since this is not true in practice [11], a smoothing filter is applied to the raw images to ensure that these assumptions are met. Such a pre-processing process causes a loss of spatial resolu-

tion [5]. On the other hand, threshold-free methodologies [12] employ erosion and dilatation morphological operators with a set of structuring elements of various sizes to detect regions exhibiting moderate activation levels and wide spatial size. However, their results are subject to the form of the structuring elements, which is determined arbitrarily.

In the current contribution, we propose a new method, denominated the regularized hypothesis testing (RHT) method. It is based on the formulation of the hypothesis testing task as a Bayesian estimation problem using a Markovian random field (MRF) to incorporate local spatial information. Firstly, mention is made of the state-of-the-art methods available to solve the problem of hypothesis testing in 2 and 3-dimensional fields. Thereafter, RHT is explained along with related theoretical considerations. The problem of parameter selection is addressed by proposing two algorithms for automatic calibration. Having laid out the new method, it is validated by experiments with simulated and real data. Finally, the results are discussed and conclusions are drawn.

Theoretical framework

The problem of testing statistical hypotheses at each element of a 2 or 3-dimensional field can be conceived as the following general problem:

Given a set $\mathcal{L}$ of sites, there is a statistic $\mathbf{T}(u)$ defined for each site $u \in \mathcal{L}$ for which one wishes to test a null hypothesis (H_0). According to this hypothesis, all elements $\mathbf{T}(u)$ are furnished by the distribution $P_0(\mathbf{T})$ (the null distribution). H_0 is assumed to be of the form:

$$H_0 = \bigcap_{u \in \mathcal{L}} H_{0u}, \tag{1}$$

where H_{0u} is a marginal null hypothesis about the probability distribution of the measurements at site u. At H_{0u}, consequently, $\mathbf{T}(u)$ is generated by P_0. In the active region, $\mathcal{A}$ is defined as the set of sites u where

H_{ou} does not hold, thus affirming the alternative hypothesis H_{1u}. The problem then is to find an estimate $\hat{\mathcal{A}}$ for te set $\mathcal{A}$ (note: $\mathcal{A}$ and $\hat{\mathcal{A}}$ may be empty).

Once a method is selected, its performance must be evaluated with standard tools. Some very common tools that will be used presently are described.

1. The false positive rate (FPR) is defined by:

$$FPR = E\left[\frac{|\mathcal{A}^c \cap \hat{\mathcal{A}}|}{|\mathcal{A}^c|}\right], \tag{2}$$

where $\mathcal{A}^c$ is the complement of the active region $\mathcal{A}$, $|\cdot|$ denotes the cardinality of a set and $E[\cdot]$ refers to the expected value of a random variable.

It is also possible to define FPR for elements that are not adjacent to the boundary of the active region:

$$FPR_r = E\left[\frac{|(D_r\mathcal{A})^c \cap \hat{\mathcal{A}}|}{|(D_r\mathcal{A})^c|}\right], \tag{3}$$

where the r-dilation D_r of $\mathcal{A}$ is defined as:

$$D_r\mathcal{A} = \left\{u \in \mathcal{L} : \min_{v \in \mathcal{A}}||u - v|| \leq r\right\}. \tag{4}$$

This measure is defined for two reasons. Firstly, the boundary of the active region is usually not well localized, since the activation level often decreases slowly as one moves away from $\mathcal{A}$. Secondly, the methods that consider the neighborhood of each element for estimating $\mathcal{A}$ likely indicated an increased number of false positives nearby to the boundary of the active region If most false positives are of this kind, they should have less impact on the measured performance than the false positives disconnected from $\mathcal{A}$.

Thus, FPR_r could be a better performance measure in such a case.

2. The true positive rate (TPR), also called sensitivity, is defined by:

$$TPR = E\left[\frac{|\mathcal{A} \cap \hat{\mathcal{A}}|}{|\mathcal{A}|}\right], \tag{5}$$

denoting the expected proportion of sites correctly estimated as the activation region.

3. The family-wise error rate (FWER) is defined by:

$$FWER = Pr(\hat{\mathcal{A}} \neq \emptyset | \mathcal{A} = \emptyset), \tag{6}$$

representing the probability of having at least one false positive, given that the activation region is empty.

4. The false discovery rate (FDR) is defined by:

$$FDR = E\left[\frac{|\mathcal{A}^c \cap \hat{\mathcal{A}}|}{|\hat{\mathcal{A}}|}\right] = \frac{W \cdot FPR}{|\mathcal{A}| \cdot FPR + TPR}, \tag{7}$$

with $W = (|\mathcal{L}| - |\mathcal{A}|)/|\mathcal{A}|$,

portraying the proportion of wrongly rejected null hypotheses in $\hat{\mathcal{A}}$.

Some of these measures can be combined. For example, a widely accepted way of characterizing the performance of a method is through the receiver operating characteristic (ROC) curve [13], which indicates, for a fixed $\mathcal{A}$, the maximum attainable TPR for any given maximum allowable FPR. For one-sided tests, the maximum TPR is generally an increasing function of the maximum allowable FPR.

To construct the curve, the true region $\mathcal{A}$ and the corresponding activation level $a = E[\mathbf{T}(u)]$, $u \in \mathcal{A}$ must be known. Then, the study of a method that depends on a parameter θ involves setting a value for θ to obtain a point on the ROC curve. The parameter acts as a cut-

off point to distinguish between the sites considered positives or negatives. Take as an example the methods based on the following computation:

$$\hat{\mathcal{A}} = \{\, u : pv(u) \le \theta \,\}, \qquad (8)$$

where $pv(u)$ depicts the p-value of $\mathbf{T}(u)$ and therefore $pv(u) = 1 - P_0(\mathbf{T}(u))$. This class of methods, denominated PW, encompasses most of the standard methods. They differ from each other only in the way the threshold θ is computed (note: all PW methods have the same ROC curve, regardless of the way θ is computed). Thus, for the application of the threshold based on a significance level α, without correction for multiple hypotheses (uncorrected method), it holds that $\theta = \alpha$. For the FWER method [6] [8] [14], with significance level α_{FWER}, it follows that:

$$\theta = 1 - P_0((P_0^M)^{-1}(1 - \alpha_{FWER})), \qquad (9)$$

where P_0^M is the distribution of $\max_{u \in L} \mathbf{T}(u)$ under H_o. Consequently, FWER can be controlled by choosing as the threshold the value located at the (1-α_{FWER}) portion of the right side of P_0^M. For an elaborate discussion on the association of FWER with the maximum statistical value, see Pantazis [15].

In the case of FDR, with a significance level α_{FDR} [16], the procedure for finding the threshold θ begins with ordering the individual p-values of sites $u \in L$ from the largest to the smallest. Accordingly, $pv(u_1) \ge pv(u_2) \ge \cdots pv(u_N)$ with $N = |\mathcal{L}|$. Let k be the index of the first site on the list, at which the p-value is less than or equal to the desired FDR proportion. Hence,

$$pv(u_k) \le \frac{k}{N}\alpha_{FDR},$$

and θ is set as:

$$\theta = pv(u_k). \qquad (10)$$

For more details on the method, consult Benjamini [17]. Alternatively, given a desired value ϵ for FPR_{max}, θ can simply be set as ϵ (e.g., $\epsilon = 10^{-5}$). If the local hypotheses H_{0u} are independent, this is equivalent to the application of the standard PW method without correction for multiple hypotheses, but with low level of significance $\alpha = \epsilon$. Since (as mentioned) the maximum TPR is an increasing function of FPR, the value of $\theta^* = \epsilon$ will be the one that maximizes the TPR while keeping FPR under control:

$$\max\ TPR\ \ subject\ to\ FPR \le \epsilon, \qquad (11)$$

where ϵ is a user-specified small positive number. Although the actual ROC curve for a given problem is unknown (because it depends on the values of $\mathbf{T}(u)$ for $u \in \mathcal{A}$), it is possible to specify the value of ϵ. Thus, the optimal PW method, according to Eq. (11) with $\theta = \theta^*$, will correspond to the UC method with $\alpha = 1 - \epsilon$. It may be applied if the field of p-values, or equivalently the field $P_0(\mathbf{T}(u))$, $u \in \mathcal{L}$, is known. Such fields can be estimated either theoretically or by non-parametric empirical means (e.g., permutation or re-sampling procedures) [18] [19]. The optimal PW method will be denoted by $PW^*(\epsilon)$.

The signal-to-noise ratio (SNR) [20] is defined as:

$$SNR = \frac{\min\limits_{u \in \mathcal{A}} \mathbf{T}(u)}{\sigma_0}, \qquad (12)$$

where σ_0 is the variance of $\mathbf{T}$ under H_0 and the minimum activation level in $\mathbf{T}$ has been arbitrarily assigned as the information signal. This represents the worst-case scenario in which the entire region $\mathcal{A}$ has a minimum value. On the other hand, the mean activation level or the amplitude could also be employed, as discussed by Welvaert and Rosseel [17] and Acosta-Franco et al. [5]. For low values of SNR, the value of TPR obtained with $PW^*(\epsilon)$ in the corresponding PW ROC curve is usually low for reasonable values of ϵ.

Consequently, PW methods have low sensitivity and the estimated $\hat{\mathcal{A}}$ is too conservative.

Hence, more sensitive methods must be developed that are able to accurately reflect the active regions, which in most cases consist of clusters of several contiguous elements of $\mathcal{L}$ and not of isolated elements. Customarily, these methods use the field $\mathbf{T}$ as a starting point for the generation of a new field $\hat{\mathbf{T}}$. The new statistic takes the spatial correlation of $\mathcal{A}$ into account, and then processes $\hat{\mathbf{T}}$ with a PW method as before. For example, $\hat{\mathbf{T}}$ may refer to the size or mass of a cluster of arc-connected elements (u) with values of $\mathbf{T}$ above a certain threshold [21].

However, this approach has some drawbacks, one being that the results depend strongly on the value of the selected threshold, and in general no principled way exists to make such a selection. Some variants of the method alleviate the problem to a certain extent by computing $\hat{\mathbf{T}}$ as a weighted combination of cluster sizes obtained with different thresholds. Another problem is the loss of interpretability, since in many cases the significance of the statistic sought is directly related to the activation level (i.e., the value of $\mathbf{T}$), and to the extension of supra-threshold clusters.

A distinct approach was developed to address these problems, being the morphology-based hypothesis testing (MBHT) method [12]. It involves the computation of $\hat{\mathbf{T}}$ as a combination of the results found when applying a set of K morphological erosions with different sizes of structuring elements to the field $\mathbf{T}$. The statistics are calculated as:

$$\mathbf{T}_k(u) = \min_{v \in W_k(u)} \mathbf{T}(v), \, for \, k = 1, \dots, K, \qquad (13)$$

where W_k (u) denotes the structuring element k (e.g., a circle of radius r_k) centered at u. Then, these values are integrated into the statistic $\mathbf{T}$:

$$\widehat{\mathbf{T}}(u) = \max\{P_{01}(\mathbf{T}_1(u)), \dots, P_{0K}(\mathbf{T}_k(u))\}, \qquad (14)$$

where P_{0K} is the null distribution of the statistic $\mathbf{T}_k$. Afterwards, the optimal PW procedure can be applied to the field $\hat{\mathbf{T}}$ as detailed above.

The results of this approach are competitive with those based on supra-threshold cluster statistics [12] and allow a clearer interpretation. Once again, the disadvantage is that the results may depend on the shape of the structuring elements, and for their selection no principled solution exists.

In the current contribution, the approach introduced is capable of overcoming these difficulties. It formulates the hypothesis testing task as a Bayesian estimation problem, with an MRF previously applied to the active regions, thus implementing a prior constraint on the spatial contiguity of $\mathcal{A}$. For the application of a Gaussian Markov random field (GMRF) for fMRI data analysis, see Mejia et al. and the references therein [3].

The new scheme proved to have better performance than PW methods, while maintaining interpretability. It also has better performance than MBHT, and does not require the selection of any particular shape for the structuring elements.

The regularized hypothesis testing method

In RHT, the hypothesis testing problem is written in terms of an image segmentation problem, which is solved by using a Bayesian estimation framework. First, the hypothesis testing formulation is explained. The procedure is laid out for calculating the prior distributions of sites and the likelihood that they belong to an active region. Subsequently, an approximation algorithm for finding the maximum a posterior probability (MAP) estimate is established. Finally, the parameter selection problem is addressed.

Hypothesis testing formulation

Hypothesis testing may be formulated as a binary segmentation problem. Accordingly, given the set of sites $\mathcal{L}$, the problem is reduced to partitioning $\mathcal{L}$ into non-overlapping cohesive regions $\mathcal{A}_0$, $\mathcal{A}_1$, such that the activation level in region $\mathcal{A}_0$ is zero, and in region $\mathcal{A}_1$ is greater than or equal to a known constant a_1. Hence, $\mathcal{A}_0 \cap \mathcal{A}_1 = \emptyset$ and $\mathcal{L} = \mathcal{A}_0 \cap \mathcal{A}_1$, where $\mathcal{A} \stackrel{def}{=} \mathcal{A}_1$ is the active region, and therefore $\mathcal{A}_0 = \mathcal{A}^C$.

Let $\mathbf{c}$ be an unknown discrete label field that identifies the partitions of $\mathcal{L}$, defining $\mathbf{c}(u) \in \mathcal{K} \stackrel{def}{=} \{0,1\}$ and $\mathbf{c}(u) = k$ if $u \in \mathcal{A}_k$, for each $u \in \mathcal{L}$. Thus, $a_{c(u)}$ depicts the activation level at site u (note: $a_0 = 0$).

Without loss of generality, the activation level in $\mathcal{A}_1$ is assumed to be equal to a_1, since this represents a worst case in terms of the desired false positive control in the estimation of $\mathcal{A}$. Considering $\mathbf{T}$ as the field formed by the statistics $\mathbf{T}(u)$ for all $u \in \mathcal{L}$, the following observation model is proposed:

$$\mathbf{T} = a_1\mathbf{c} + n, \tag{15}$$

where $\boldsymbol{n}$ is a noise field showing distribution $P_n(\boldsymbol{n})$, with:

$$P_\mathbf{n}(\mathbf{n}) = \frac{1}{Z_\mathbf{n}}\exp\left(-U_\mathbf{n}(\mathbf{n})\right), \tag{16}$$

where Z_n is a normalization constant and $U_n(\mathbf{n})$ is a so-called energy function. Then, the likelihood $P(\mathbf{T}|\mathbf{c})$ is furnished as:

$$\begin{aligned} P(\mathbf{T}|\mathbf{c}) &\stackrel{def}{=} \frac{1}{Z_{\mathbf{T}|\mathbf{c}}}\exp\left(-U_{\mathbf{T}|\mathbf{c}}(\mathbf{c})\right) \\ &= P_\mathbf{n}(\mathbf{T} - a_1\mathbf{c}), \end{aligned} \tag{17}$$

where $Z_{(\mathbf{T}|\mathbf{c})}$ is a normalization constant. It is important to find an estimate that considers the spatial correlation in both the noise field $\mathbf{n}$ and the label field $\mathbf{c}$.

As the real statistic $\mathbf{T}$ is obtained through a regression analysis of data provided by a model, it could take on a positive or negative value (depending on the choice of the statistic), resulting in positive or negative differences– activation regions– of the data with respect to the model. Zero-value regions would still portray sites where there is no activation, and the distribution of values of field $\mathbf{n}$ would be determined by the selection of the statistic $\mathbf{T}$. Hereafter, a noise model is presented as a GMRF. In later sections, we remove this assumption by introducing a procedure for standardizing the noise field in order to map noise fields ranging from arbitrary distributions to standard normal distributions.

Noise field model

For the noise field $\mathbf{n}$, a GMRF model was employed, taking the form of Eq. (16) with:

$$U_\mathbf{n}(\mathbf{n}) = \frac{1}{2}\gamma\sum_{u \in \mathcal{L}}\mathbf{n}(u)^2 + \tau_1\sum_{\langle u,v\rangle}(\mathbf{n}(u) - \mathbf{n}(v))^2, \tag{18}$$

where $\gamma > 0$ is a scale parameter, $\mathbf{n}(u)$ is the value of the field $\mathbf{n}$ in site $\mathbf{u}$, and $\mathbf{n}(v)$ is the value of the field n in a neighboring site v. The parameter $\tau_1 \geq 0$ is related to the spatial correlation of the field $\mathbf{n}$ and the second term added to the first is taken from all pairs of neighboring sites $\langle u, v\rangle$ in the image.

When $\tau_1 = 0$ and $\gamma = 1$, the formulation is reduced to a Gaussian white noise model with zero mean and unit variance (the "Parameter Selection" section explains other types of noise fields). With such a model, the likelihood takes the form of Eq. (17), being:

$$\begin{aligned} U_{\mathbf{T}|\mathbf{c}}(\mathbf{c}) = \frac{1}{2}\gamma\sum_{u \in \mathcal{L}}&\mathbf{c}(u)(\mathbf{T}(u) - a_1\mathbf{c}(u))^2 \\ &+ (1 - \mathbf{c}(u))\mathbf{T}(u)^2 \\ &+ \tau_1\sum_{\langle u,v\rangle}(\mathbf{T}(u) \\ &- a_1\mathbf{c}(u) - \mathbf{T}(v) \\ &+ a_1\mathbf{c}(v))^2. \end{aligned} \tag{19}$$

In the above equation, $\mathbf{c}(u)$ is used as an indicator function. Accordingly, if its value is 1 (i.e., u ∈ $\mathcal{A}_1$), the term $(1-\mathbf{c}(u))\mathbf{T}(u)^2$ becomes zero, simplifying Eq. (19) to an expression that is equivalent to substituting $\mathbf{n}(u)$ for $\mathbf{T}(u) - a_1\mathbf{c}(u)$. In Eq. (18), the activation level in $\mathbf{T}(u)$ is eliminated, keeping only the noise component, as in the observation model proposed in (15).

Moreover, if $\mathbf{c}(u) = 0$ (i.e., u ∈ $\mathcal{A}_0$), the term $\mathbf{c}(u)(\mathbf{T}(u) - a_1\mathbf{c}(u))^2$ becomes zero and the term $(\mathbf{T}(u) - a_1\mathbf{c}(u))$ in the second summation is simplified to $\mathbf{T}(u)$. This is equivalent to substituting $\mathbf{n}(u)$ for $\mathbf{T}(u)$ in Eq. (18), since $\mathbf{T}(u)$ already corresponds to purely noise.

Eq. (19) may be rewritten as a quadratic function of a vector-valued field $\mathbf{b}$ defined as $\mathbf{b}_1(u) = \mathbf{c}(u)$ and $\mathbf{b}_0(u) = 1 - \mathbf{c}(u)$,

$$U_{\mathbf{T}|\mathbf{b}}(\mathbf{b}) = \frac{1}{2}\gamma \sum_{u \in \mathcal{L}} \sum_{k \in \mathcal{K}} (\mathbf{T}(u) - a_k)^2 \mathbf{b}_k^2(u)$$
$$+ \tau_1 \sum_{\langle u,v \rangle} \sum_{i,j \in \mathcal{K}} (\mathbf{T}(u) \qquad (20)$$
$$- a_i - \mathbf{T}(v)$$
$$+ a_j)^2 \mathbf{b}_i(u)\mathbf{b}_j(v),$$

with $a_0 = 0$ and:

$$\sum_{k \in \mathcal{K}} \mathbf{b}_k(u) = 1, \forall u \in \mathcal{L}, \qquad (21)$$

$$\mathbf{b}_k(u) \in \{0,1\}, \forall u \in \mathcal{L}, k \in \mathcal{K}. \qquad (22)$$

Bayesian formulation of the hypothesis testing problem

For field $\mathbf{b}$, we propose the previous application of an MRF model [22] [23]. In order to introduce the prior constraint that the active regions are spatially cohesive, $\mathbf{b}$ is modeled with Ising potentials, furnishing the following distribution:

$$P(\mathbf{b}) = \frac{1}{Z_{\mathbf{b}}} \exp(-U_{\mathbf{b}}(\mathbf{b})), \qquad (23)$$

$$U_{\mathbf{b}}(\mathbf{b}) = \tau_2 \sum_{\langle u,v \rangle} \|\mathbf{b}(u) - \mathbf{b}(v)\|^2, \qquad (24)$$

where $\tau_2 \geq 0$ is a parameter that controls the granularity of field $\mathbf{b}$.

With this equation and the Gauss-Markov model (16) and (18) for noise, it is possible to estimate field $\mathbf{b}$ by employing a Bayesian formulation, affording the posterior probability distribution:

$$P(\mathbf{b}|\mathbf{T}) \propto P(\mathbf{T}|\mathbf{b})P(\mathbf{b}), \qquad (25)$$

where the likelihood is:

$$P(\mathbf{T}|\mathbf{b}) \propto \exp(-U_{\mathbf{T}|\mathbf{b}}(\mathbf{b})), \qquad (26)$$

with $U_{\mathrm{T|b}}$ given by (20) and field $\mathbf{b}$ satisfying (21)-(22). Finally, the posterior distribution can be written as:

$$P(\mathbf{b}|\mathbf{T}) \propto \exp(-U_{\mathbf{b}|\mathbf{T}}(\mathbf{b})), \qquad (27)$$

$$U_{\mathbf{b}|\mathbf{T}}(\mathbf{b}) = U_{\mathbf{T}|\mathbf{b}}(\mathbf{b}) + U_{\mathbf{b}}(\mathbf{b}), \qquad (28)$$

where $U_{\mathrm{T|b}}(\mathbf{b})$ and $U_{\mathbf{b}}(\mathbf{b})$ are furnished by (20) and (24) respectively.

The MAP estimate for field $\mathbf{b}$ is obtained by minimizing (28), subject to the constraints (21)-(22). This is a combinatorial optimization problem that is, in general, very difficult to solve. Several methods have been proposed, such as the iterated conditional modes (ICM) algorithm [24], stochastic relaxation [25], graph cuts procedure [26], etc.

Here, because of the form of the cost function that includes the noise correlation term, we prefer option [22], which is based on the relaxation of constraint (22). Hence, a new formulation is made:

$$\mathbf{b}_k(u) \geq 0, \forall k, u. \qquad (29)$$

Dividing (28) by γ and utilizing $v \overset{def}{=} \tau_1/\gamma$ and $\lambda \overset{def}{=} \tau_2/\gamma$, the functional is minimalized as:

$$U(\mathbf{b}; v, a_1, \lambda) = \frac{1}{2} \sum_{u \in \mathcal{L}} \sum_{k \in \mathcal{K}} (\mathbf{T}(u) - a_k)^2 \mathbf{b}_k^2(u)$$
$$+ v \sum_{\langle u,v \rangle} \sum_{i,j \in \mathcal{K}} \big(\mathbf{T}(u) - a_i$$
$$- \mathbf{T}(v) + a_j\big)^2 \mathbf{b}_i(u)\mathbf{b}_j(v) \qquad (30)$$
$$+ \lambda \sum_{\langle u,v \rangle} \|\mathbf{b}(u) - \mathbf{b}(v)\|^2.$$

If the noise has a mean of zero, then $a_0 = 0$. The resulting optimization problem is:

$$\text{Find } \mathbf{b}^*(v, a_1, \lambda)$$
$$= \arg \min_{\mathbf{b}} \ U(\mathbf{b}; v, a_1, \lambda), \qquad (31)$$

$$s.t: \sum_{k \in \mathcal{K}} \mathbf{b}_k(u) = 1, \forall u \in \mathcal{L}, \qquad (32)$$

$$\mathbf{b}_k(u) \geq 0, \forall u \in \mathcal{L}, k \in \mathcal{K}. \qquad (33)$$

This is a quadratic minimization problem, subject to linear constraints, which can be solved efficiently by employing a projected gradient descent method. Consequently, each $\mathbf{b}(u)$ (in order to simplify the notation we remove the dependency of v, a_1, λ) may now be interpreted as an approximation for the posterior marginal distribution of the indicator functions $1_{\mathcal{A}k(u)}$.

To obtain the latter it is possible to set $1_{\mathcal{A}k(u)} = 1$ if $\mathbf{b}_{k(u)} > \mathbf{b}_{j(u)}$ for all $j \neq k$. Thus, the estimation of the active region is:

$$\hat{\mathcal{A}} = \{ u \in \mathcal{L} \mid \mathbf{b}_1(u) > 0.5 \}. \qquad (34)$$

Parameter selection

A general method can now be introduced to allow the model (31)-(33) to be used for any type of noise. The parameters a_1, v, λ will be able to be automatically selected by controlling the false positive proportion while maximizing the performance of the method with respect to the TPR.

Standardization of the noise field

The derivation of functional (31) is based on the observation model (15) with the Gauss-Markov model (16)-(18) for noise. However, the model is not always valid in practice, since the **T** field may have a different distribution. The first step, therefore, is to generate a transformation capable of mapping the arbitrary distribution to a standard one. The proposed method is based on two properties:

- If x is a random variable and Q its corresponding *cumulative distribution function* (*CDF*), then the distribution function for the random variable $Q(x)$ is uniform [27] [28].
- If u is a uniform random variable and G the *CDF* of the normal distribution, then the distribution function for the random variable $G^{-1}(u)$ is normal [29].

Combining both properties, it follows that for any random variable x, the random variable $G^{-1}(Q(x))$ is normal. Consequently, the only requirement for transforming a random variable x into a normal variable is to know the corresponding *CDF* $Q(x)$ or to estimate it by obtaining the empirical CDF (ECDF) $\hat{Q}(x)$ from samples of the particular distribution.

Assuming that random field samples $\mathbf{T}_0$ can be generated under H_0, it is possible to estimate the corresponding null distribution $P_0(\cdot)$. Given a field **T** derived from the observations (e.g., the *F*-test for the generalized linear model (GLM) computed at each voxel), the sample can be mapped to a standard normal distribution Φ:

$$\Phi(x) = \frac{1}{2}\left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right), \qquad (35)$$

where $erf(y)$ denotes the error function [30], the probability that a random variable Z normally distributed with mean $\mu = 0$ and variance $\sigma = \frac{1}{2}$ falls in the range [-y, y]. The field $\hat{\mathbf{T}}$, with a standard normal distribution, is afforded by the following transformation:

$$\hat{\mathbf{T}}(u) = \sqrt{2}erf^{-1}\big(2P_0\big(\mathbf{T}(u)\big) - 1\big), \forall u \in \mathcal{L}. \tag{36}$$

Hyper-parameter calibration

The hyper-parameters that control the behavior of the method are the spatial correlation parameter v, the minimum activation level a_1, and the regularization parameter λ. The general idea is to estimate v from the observations, and subsequently compute a_1, λ from the estimated parameter v and a parameter ϵ representing the level of FPR control.

Estimation of parameter v: The parameter $v = \tau_1/\gamma$ in (31) controls the noise spatial correlation and can be estimated by minimizing the negative logarithm of the pseudo-likelihood of samples of the field $\mathbf{T}_0$ generated under H_0 (i.e., samples of the noise field $\mathbf{n}$).

The pseudo-likelihood is defined as the product of the conditional distributions for $\mathbf{n}(u)$ at each pixel u, given the values at its neighboring fields [31]. Using Eqs. (17)-(18), the result is:

$$\begin{aligned} L(\gamma,\tau_1) &= -\log \prod_{u\in\mathcal{L}} P(\mathbf{n}(u)|\mathbf{n}(v), v \in \mathcal{N}_u, \gamma, \tau_1) \\ &= \left(\tfrac{\gamma}{2}+\tau_1|\mathcal{N}_u|\right)p - 2\tau_1 q + \tfrac{\tau_1^2 r}{\frac{\gamma}{2}+\tau_1|\mathcal{N}_u|} - \\ &\quad \tfrac{|\mathcal{L}|}{2}\log\left(\tfrac{\gamma}{2}+\tau_1|\mathcal{N}_u|\right) + t, \end{aligned} \tag{37}$$

where t is a constant, $\mathcal{N}$ is the neighborhood of pixel u and:

$$p = \sum_{u\in\mathcal{L}} \mathbf{n}(u)^2 \tag{38}$$

$$q = \sum_{u\in\mathcal{L}} \mathbf{n}(u)\mathbf{s}(u) \tag{39}$$

$$r = \sum_{u\in\mathcal{L}} \mathbf{s}(u)^2 \tag{40}$$

$$with\ \mathbf{s}(u) = \sum_{v\in\mathcal{N}_u} \mathbf{n}(v). \tag{41}$$

The minimization of Eq. (37) yields a closed formula for the estimation of the parameters:

$$\tau_1 = \frac{q|\mathcal{L}|}{2(pr-q^2)}, \tag{42}$$

$$\gamma = 2\left(\tfrac{r}{q} - |\mathcal{N}_u|\right)\tau_1, \tag{43}$$

$$\hat{v} = \tfrac{\tau_1}{\gamma}. \tag{44}$$

Calibration of parameters a_1, λ: For a given value ϵ for FPR control and a fixed $v = \hat{v}$ calculated with Eq. (44), the present proposal is to find the optimal parameter settings a_1^*, λ^* as the maximizers of the TPR, subject to FPR $\leq \epsilon$. However, the solution of such an optimization problem is very difficult to obtain because both TPR and FPR depend on the unknown region A. To find an approximate solution, the following observations are made:

1. Instead of $FPR \leq \epsilon$, we use $FPR_2 \leq \epsilon$, Eq. (3), since the precise location of the boundary of the active region $\mathcal{A}$ is uncertain (as aforementioned) and in some sense arbitrary, considering that activation generally has a gradual variation in the field.

2. According to our experimental work (see Fig. 2 and Experiment 1 below), the present method has the advantage that false positives do not extend, in a significant way, more than a couple of pixels away from this boundary. This is due to boundary effects close to the active region. For a given fixed a_1 and λ, therefore, it is possible to estimate the constraint $FPR_2 \leq \epsilon$ by FPR^0 (i.e., the FPR computed in fields generated under H_0), which is depicted by $FPR^0(a_1, \lambda)$.

3. The calculation can be further simplified based on the observation that TPR is an increasing function of FPR (i.e., the usual shape of the ROC curves). Consequently, for a fixed λ:

$$\hat{a}_1(\lambda) = \arg \max_{a_1} \ TPR(a_1, \lambda), \qquad (45)$$

which is equivalent to finding $a_1(\lambda)$, such that $FPR^0(a_1(\lambda), \lambda) = \epsilon$. Given that a $\hat{a}_1$ is obtained from fields produced under H_0 and these fields are standardized as described above with the noise model (18), the values of a a_1 as a function of v, λ, ϵ do not depend on the data. Hence, it is possible to pre-compute them based on fields generated by utilizing model (16) with U afforded by (18), being then the Gibbs sampler algorithm [25].

4. Since the TPR depends on the unknown $\mathcal{A}$, we propose the approximation of TPR for each value of λ by the average $\overline{TPR}(\lambda)$. The latter value results from running the method with the value $\hat{a}_1(\lambda)$ for a $\hat{a}_1$ over a finite set of values $\hat{S}$, which is a discretized version of the collection of activation values S in a continuous interval. We use a fixed synthetic active region (a circle):

$$\overline{TPR}(a_1, \lambda; v, \epsilon) =$$
$$\int_{a \in S} TPR(a, a_1, \lambda; v, \epsilon) P(a) \mathrm{d}a \approx \qquad (46)$$
$$\frac{1}{|\hat{S}|} \sum_a TPR(a, \hat{a}_1(\lambda; v, \epsilon), \lambda; v, \epsilon),$$

where $P(a)$ is taken as the uniform distribution in the set S. With these simplifications, the parameter estimation algorithm simply consists of sweeping the plausible λ values in a given set Λ (a discretization of an interval $[0, \lambda_{max}]$), followed by finding, for each λ, the value $\hat{a}_1(\lambda)$ (such that FPR $(\hat{a}_1(\lambda), \lambda) = \epsilon$) and the corresponding $(\overline{TPR})(\lambda)$, as explained above. The final λ^* is then found as the maximizer of $(\overline{TPR})(\lambda)$, affording $a_1^* = \hat{a}_1(\lambda^*)$. This is summarized in Algorithm 1.

The optimal parameters a_1^*, λ^* depend on the data only through the estimated parameter v and the desired FPR^0 control ϵ. Thus, the optimal values may be pre-computed and stored in a table. When a new data set arrives, one only needs to standardize the noise, estimate v, read a_1^*, λ^* from the table for the desired FPR control, and minimize U ($\mathbf{b}$; v, a_1^*, λ^*) furnished by Eq. (30). The tables are represented as false color images in Figure 1.

The final algorithm for estimating the active region $\hat{\mathcal{A}}$ for a given data set is summarized in Algorithm 2.

Algorithm 1: Calibration parameter algorithm (CPA)		
	function CPA(v, ϵ)	
1	**Input**	Estimated value of v from (44); desired control ϵ for the FPR; search interval Λ for λ.
	Output	Estimated hyperparameter values a_1^*, λ^*
2	**Begin**	
3	For all $\lambda \in \Lambda$ **do**	
4	Compute $\hat{a}_1(\lambda; v, \epsilon)$ that solves $FPR(a_1, \lambda; v) = \epsilon$;	
5	Compute $\overline{TPR}(\hat{a}_1(\lambda; v, \epsilon), \lambda; v, \epsilon)$ by using (46);	
6	**End**	
7	Compute $\lambda^* = \arg \max_{\lambda \in \Lambda} \overline{TPR}(\hat{a}_1(\lambda; v, \epsilon), \lambda; v, \epsilon)$	
8	Compute $a_1^* = \hat{a}_1(\lambda^*(v, \epsilon); v, \epsilon)$;	
9	**Return** a_1^*, λ^*	
10	**End**	

Algorithm 2: Regularized hypothesis testing (RHT) algorithm		
	function RHT $(\mathbf{T}, \hat{\epsilon}, a_1^*(v, \epsilon), \lambda^*(v, \epsilon))$	
1	**Input**	Observed field $\mathbf{T}$, parameters $a_1^*(v, \epsilon)$, $\lambda^*(v, \epsilon)$ computed with Algorithm 1 and the desired control for the FPR
	Output	Estimated active region $\hat{\mathcal{A}}$
2	**Begin**	
3	Compute $\hat{\mathbf{T}}$ with (36);	
4	Compute $\hat{v}$ with (44) using the field $\hat{\mathbf{T}}$;	
5	Compute $\hat{a}_1 = a_1^*(\hat{v}, \hat{\epsilon})$ by interpolating $a_1^*(v, \epsilon)$ from a pre-computed table;	
6	Compute $\hat{\lambda} = \lambda^*(\hat{v}, \hat{\epsilon})$ by interpolating $\lambda^*(v, \epsilon)$ from a pre-computed table;	
7	Compute $b^*(\hat{v}, \hat{a}_1, \hat{\lambda})$ by solving (31)-(33);	
8	Compute $\hat{\mathcal{A}}$ with (34);	
9	**Return** $\hat{\mathcal{A}}$	
10	**End**	
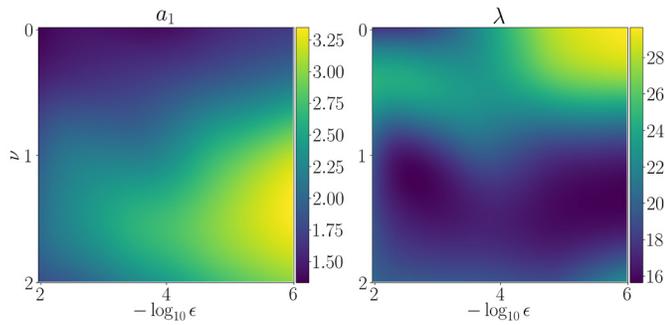
FIGURE 1. Estimation of parameters a_1, λ as functions depending on $(v, -log_{10}\epsilon)$ for $v \in$ [0, 2] and $\epsilon \in$ [10^6, 10^{-2}].

MATERIALS AND METHODS

Data description

To study the behavior of the present method, we include various experiments based on both synthetic and real data.

Synthetic data: The synthetic data was generated dynamically and consisted of two components. Firstly, $S_0 = \{n_1, n_2, ..., n_{40}\}$ is a set of noise fields of a regular lattice $\mathcal{L}$ of 50×50 pixels, without any activation region. That is, the images were produced under H_0, using Eqs. (16) and (18) with $n(u)$ as random independent variables showing standard Gaussian distribution. Parameter $\gamma = 1$ and v were to be specified in each experiment. Secondly, $S_1 = \{c_1, c_2, ..., c_{40}\}$ consists of 40 binary labeled fields of a regular lattice of 50×50 pixels, representing indicator functions for activation regions of different shapes (see Figure 2). Then, the simulated observed fields are generated by the following model:

$$\hat{a}_1(\lambda) = \arg \max_{a_1} TPR(a_1, \lambda), \qquad (47)$$

where a denotes the activation level and its value is specified in each experiment.

Real data: Experiments with real data were based on the auditory dataset [32]. The data corresponded to 96 volumes of a single subject (each volume composed of

64×64×64 voxels of 1×1×3 mm), which were acquired in blocks of 6 volumes. Since the repetition time between scanning was set to 7 seconds, there were a total of 16 blocks of 42 seconds (although due to the effects related to T1, the first two blocks were discarded). The sequence of volumes alternated between blocks of rest and stimulation, starting with rest. Auditory stimulation consisted of two-syllable words presented binaurally at a rate of 60 per minute.
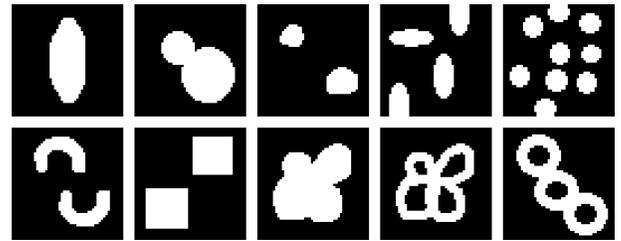
FIGURE 2. Dataset: Examples of synthetic active regions.

Localization of false positives for the RHT algorithm

This experiment was designed to test the assumption that in the proposed method the false positive errors are concentrated close to the boundary of the active region. Consequently, FPR_2 can be well approximated by FPR_0 (i.e., FPR computed over images produced under H_0).

A total of 1000 independent runs were conducted for the procedure. Briefly, set S_0 was generated with parameter $v = 0.75$, a value estimated from fMRI images by utilizing Eq. (44). Fields $\mathbf{T_1}, \mathbf{T_2}, ..., \mathbf{T_{40}}$ were obtained by using model (47), sets S_0 and S_1, and randomly selected activation level α in the interval [2] [4].

Algorithm 2 was applied for levels of FPR control $\epsilon \in$ {0.001,0.0001,0.00001,0.000001} with the optimally calibrated parameters, computing the average value of FPR_2 for the set S_1 and FPR for the set S_0. The results are shown in Table 1. As can be appreciated, by calibrating the parameters of the method to control FPR_0, the appropriate control over FPR_2 is also achieved.

Performance of the method

The first set of the following experiments was designed to make a comparison of RHT to FWER [8] and MBHT [12], the latter two being the state-of-the-art methods that take the spatial expanse of the active region into account. Accordingly, the set S_0 was produced with parameter $v \in \{0, 0.5, 1, 1.5\}$. A total of 1000 runs were performed, randomly selecting the active regions $\mathcal{A}$ from $\mathcal{S}_1$ with an activation level $a \in \{0.0, 0.25, 0.5, \cdots, 5\}$. The average TPR was computed. The RHT method was carried out by employing Algorithm 2 and a level of false positive control of $\hat{\epsilon} = 0.000001$.

Based on the results from each method, RHT proved capable of providing better TPR values than the other two algorithms for most activation levels (Figure 3). The improvement is likely due to the consideration in the new model of the spatial cohesion of the activation regions (controlled by the regularization parameter λ) as well as the spatial correlation in the noise field (controlled by the parameter v). Contrarily, the rest of the algorithms assume that these two components of statistic **T** are formed by independent variables.

Experiments with fMRI data

In the second set of experiments, Algorithm 2 of the proposed method was applied to the real fMRI data described at the beginning of this section. RHT was not applied directly to the original data, but rather to a field **T** (an *F*-test field computed from the data). The first step was to standardize the field in order to obtain $\hat{\mathbf{T}}$ with Eq. (36), which is the input of the segmentation algorithm. Hence, it was necessary to calculate the ECDF from the data. Here after, some details about the aforementioned steps are explained.

Data pre-processing. The aim of the pre-processing was to remove artifacts in the data as well as to prepare the data to maximize the statistical analysis. Here we use spatial pre-processing provided in the script auditory_spm12_batch.m, which implies realignment,
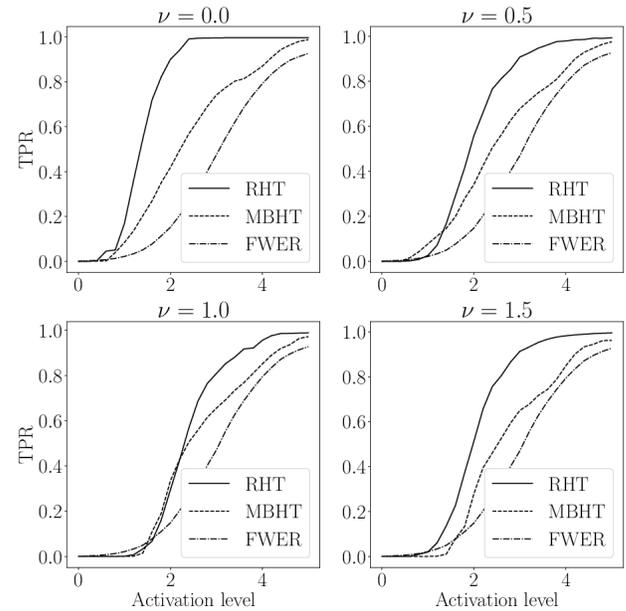
FIGURE 3. Performance comparison of RHT with the punctual FWER and MBHT methods for different noise spatial correlation v and ϵ = 0.000001.

co-registration, segmentation and normalization. Although the original script (available online at *http://www.fil.ion.ucl.ac.uk/spm/data/auditory/*) includes a smoothing step to ensure that some assumptions about noise distribution are fulfilled [11] [18], RHT omits this step because it is capable of handling different noise distributions. Actually, the inclusion of the step would not be beneficial. The data for these calculations was processed on SPM software version 12 (available at *http://www.fil.ion.ucl.ac.uk/spm/*) and MATLAB 2018a. A slice of the pre-processed data was selected to perform the following experiments.

F-test field. After the pre-processing stage, the statistical analysis was carried out to determine the active voxels that correspond to a given stimulus. To obtain the active regions, a voxel-wise analysis is typically conducted by fitting models to a single voxel time course. The data at each voxel is modeled in a univariate way with a linear model:

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t, \tag{48}$$

TABLE 1. Experiment of the FPR control at different levels, utilizing a synthetic dataset. First column, level of FPR control; second column, average FPR for images generated under H_0; third column, average FPR_2 for synthetic active regions.

ϵ	FPR^0	FPR^2
0.01	0.0042504	0.0051335
0.001	0.0001340	0.0001714
0.0001	0.0000436	0.0000436
0.00001	0.0000044	0.0000050
0.000001	0.0000004	0.0000004

where y_t, the dependent variable, is a vector formed by the intensity values at each time point. β_0 and β_1 are the intercept and the slope of the linear model, respectively. Meanwhile, ϵ_t is the error term in the model fitting, which captures factors other than x_t, such as signal noise, capable of affecting y_t. The explanatory variable x_t corresponds to the model of neural activity. It is also a vector comprised of entries depicting a real value at each time point:

$$x_t = d_t * h_\mathrm{t} \qquad (49)$$

where h_t stands for the hemodynamic response function and $d_t \epsilon \{0,1\}$ is an impulse train that indicates whether the stimulus was present at time t. Then, the neural activity is modeled as a convolution,* in which the *hemodynamic response function* acts as a filter.

The F-test represents the ratio between the variance described in a reduced model $y_t = \beta_0 + \epsilon_t$ (without any additional effect) and the full model (48)-(49) that includes the effect of interest.

Finally, the F-test was computed for every voxel in the volume. The resulting field was the input for the RHT algorithm, calculated by using the script auditory_spm12_batch.m with the default parameters.

Null distribution. Before the segmentation step of the RHT algorithm, it was necessary to normalize the F-test field and thus to estimate the null distribution. More specifically, a sample of the F-test was obtained under H_0, allowing for the computation of the corresponding ECDF. This was carried out by permuting the order of the stimulus labels $\boldsymbol{d}(\boldsymbol{t})$ for the volumes (see [33] for more details), and by calculating the F-test field for each permutation, as explained in the previous section. After completion of these procedures, it was possible to apply Eq. (36) to transform the original F-test field (i.e., the original order of the labels $\boldsymbol{d}(\boldsymbol{t})$, without permutation) to one that follows a standard normal distribution. Finally, for the $\hat{\mathbf{T}}$ fields generated under H_0, the value of v was estimated by utilizing Eq. (44), finding $v = 0.75$.

RHT algorithm robustness with respect to SNR

In order to investigate the stability of the present method, the algorithm was tested by modifying the SNR in the data from which the activation region would be established. Accordingly, blocks of observations were eliminated from the full experiment. As aforementioned, the first two of the 16 original blocks were eliminated due to effects related to T1. Hence, 14 blocks (84 volumes) were taken initially to perform the procedure described earlier in this section to identify the activation regions. Subsequently, the number of blocks was reduced by two, taking 12 blocks (72 volumes) and performing the procedure again to determine the activation regions, and so on until reaching 4 blocks.

In each case, a tolerance for false positives to $\epsilon = 0.0001$ was established and the results of RHT were compared to those obtained with MBHT and FWER using the same number of blocks. It can be verified by visual inspection (Figure 4) that the three methods– RHT, MBHT and FWER– detected activation regions corresponding to the primary auditory cortex, located at the upper sides of the temporal lobes, specifically on the transverse temporal gyri [34] [35].
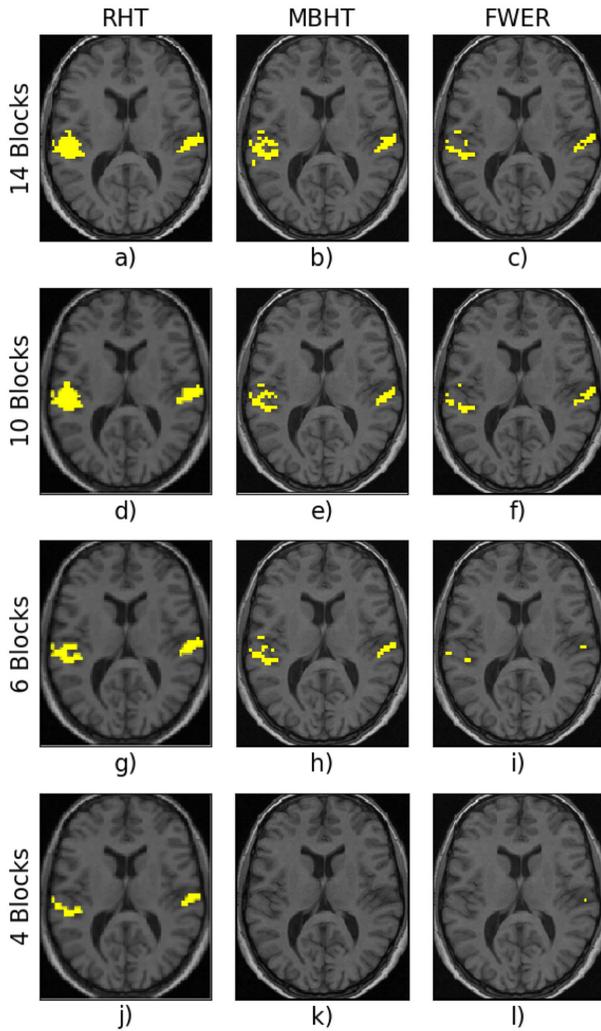
FIGURE 4. Performance comparison of the RHT, MBHT and FWER by modifying the SNR in the data through the elimination of blocks.

The MBHT and RHT methods afforded better performance and robustness for the reduction of information in most cases. MBHT and FWER displayed less accuracy when reducing the signal information for ϵ = 0.00001. While FWER exhibited difficulties in the identification of activation regions with 6 and 4 blocks (barely finding any region), RHT and MBHT gave better performance. However, when the number of blocks was decreased to only 4, MBHT was not able to detect any activation region, while RHT still had a proficient outcome. RHT outperformed the other models two by successfully recovering both regions.

Degree of false positive control. To avoid having to specify a particular value for ϵ, it is possible to estimate $\mathcal{A}$ for a set $\mathcal{S}$ of ϵ values and produce a map in which the degree of false positive control (log ϵ for each $\hat{\mathcal{A}}$ (ϵ) = RHT(**T**, $\hat{\epsilon}$, a_1^* (v, ϵ), λ^* (v, ϵ)) is color coded. The result is very similar to the p-value maps that are created for the classic PW method. Specifically, we define the degree of false positive control at site u as:

$$max_{\epsilon \in \mathcal{S}}[(-\log \epsilon)\mathbf{1}_{\hat{\mathcal{A}}(\epsilon)}(u)], \qquad (50)$$

where $\mathbf{1}_{\hat{\mathcal{A}}(\epsilon)}$ (u) is the indicator function for the set A $\hat{\mathcal{A}}(\epsilon)$. An example of this type of map is shown in Figure 5, with $\mathcal{S}$ = {0.01,0.001,0.0001,0.00001}. The map portrays important differences between the first and second levels, but in the last levels the detected regions are more similar to each other. The slight variations exist only at the border of the active region.

RESULTS AND DISCUSSION

We herein describe a new method, RHT, for detecting active regions in random fields. The focus is on applications for neuroimaging. With the present method, the expected TPR is maximized while the false positives are kept under control by specifying an upper bound ϵ on FPR^0 (the FPR under H_0). By making ϵ small enough, there is an effective correction for multiple hypotheses since the total number of false positives in the entire field is controlled. We demonstrated experimentally that most of the false positives produced by
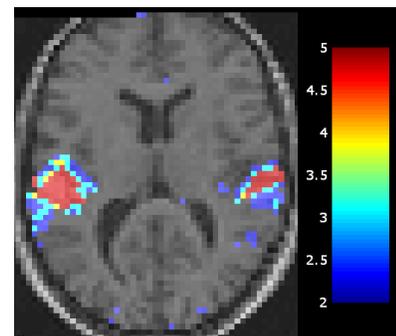
FIGURE 5. Degree of false positive control over fMRI data described in the text.

RHT are confined to the vicinity of the boundary of the active region. Thus, ϵ also bounds the FPR in the data outside of this boundary (i.e., the FPR_2).

CONCLUSIONS

The main contributions of the present model are the following:

1. A Bayesian formulation of the hypothesis testing problem in random fields was reduced to an image or volume segmentation problem, for which the maximum a posteriori estimate could be calculated.
2. A new Markovian random field model for correlated noise was implemented, from which the appropriate prior distribution could be computed.
3. A method for estimating the hyper-parameters of the model was conceived, involving two main factors: a) the application of a closed formula for the noise correlation parameter v based on the maximization of the pseudo-likelihood of the data, and b) a pre-calculated lookup table (independent of the data) for the λ and a_1 parameters. This method makes the whole procedure computationally efficient, since the only thing needed for its application is a way of generating sample images from the null distribution. Such samples can be obtained, for example, by using permutation procedures.
4. To avoid having to specify a particular value for the ϵ parameter, we demonstrated how to present a family of solutions for different values of ϵ in a single image (the DFPC map), similar to the classic p-value maps.
5. The performance of the method was validated with synthetic and real data (fMRI images). In both cases, RHT provided an improvement in the TPR (while maintaining the FPR under control) compared to the competitive state-of-the-art methods (MBHT and FWER). For the fMRI data, RHT displayed the best sensitivity, which was particularly high for low SNR.

For simplicity, in the current analysis we focused on two-dimensional data. However, it is possible to directly extend the method to 3D simply by considering an extended neighborhood (e.g., 6 or 26 neighbors for each voxel) in the prior MRF models for the active region and the noise, at the expense of an increased computational complexity.

Although the example application here corresponds to fMRI, the RHT method may be applied to any situation involving the testing of a field of local hypotheses, such as the ones that are common in neuroimaging, remote sensing, and so on.

ABBREVIATIONS

CDF, cumulative distribution function
ECDF, empirical cumulative distribution function
FDR, false discovery rate
FPR, false positive rate
fMRI, functional magnetic resonance imaging
FWER, family-wise error rate
GMRF, Gaussian Markov random fields
H_o, null hypothesis
MBAT, morphology-based hypothesis testing
MAP, maximum a posterior probability
MRF, Markovian random field
PW, pointwise method
RHT, regularized hypothesis testing
ROC, receiver operating characteristic
SNR, signal-to-noise ratio
TPR, true positive rate

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS STATEMENT

O.D. and J.M. conceived the method and the experiments, the latter conducted by D.A. O.D., D.A. and J.M. analyzed the results. All authors reviewed and approved the final version of the manuscript.

REFERENCES

[1] Cole MW, Ito T, Bassett DS, Schultz DH. Activity flow over resting-state networks shapes cognitive task activations. Nature Neuroscience. 2016 October; 19(0): p. 1718-1726, **doi: 10.1038/nn.4406.**

[2] Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, Munafo MR, et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience. 2017 January; 18(0): p. 115-126, **doi: 10.1038/nrn.2016.167.**

[3] Mejia A, Yue YR, Bolin D, Lindren F, Lindquist MA. A Bayesian General Linear Modeling Approach to Cortical Surface fMRI Data Analysis. Journal of the American Statistical Association. 2019 April; 0(0): p. 1-20, **doi: 10.1080/01621459.2019.1611582.**

[4] Flores-Leal M, Sacristán-Rock E, Jiménez-Ángeles L, Leehan JA. Primed low frequency transcranial magnetic stimulation effects on smoking cue-induced craving. Revista mexicana de ingeniería biomédica. 2016 April; 37(1): p. 39-48, **doi: 10.17488/RMIB.37.1.3.**

[5] Acosta-Franco JA, Saiffe-Farías AF, Gómez-Velázquez FR, González-Garrido AA, Romo-Vázquez RC. Activación Hemisférica Cerebral en Adultos Jóvenes Mientras Ejecutan Tareas Ortográficas Utilizando fMRI. Revista Mexicana de Ingeniería Biomédica. 2019 August; 40(2): p. 1-10, **doi: 10.17488/RMIB.40.2.5.**

[6] Lindquist MA, Mejia A. Zen and the Art of Multiple Comparisons. Psychosomatic medicine. 2015 February; 77(2): p. 114-125, **doi: 10.1097/PSY.0000000000000148.**

[7] Woo CW, Krishnan A, Wage TD. Woo, C.-W., Krishnan, A., & Wager, T. D. (2014). Cluster-extent based thresholding in fMRI analyses: Pitfalls and recommendations. NeuroImage. 2013 December; 91(0): p. 412-419, **doi: 10.1016/j.neuroimage.2013.12.058.**

[8] Abdi H. The Bonferonni and Sidak corrections for multiple comparisons. In Salkind NJ, editor. Encyclopedia of Measurement and Statistics. Thousand Oaks, CA, USA: Sage Publications; 2007. p. 103-107, **doi: 10.4135/9781412952644.**

[9] Hayasakaa S, Nichols TE. Validating cluster size inference: random field and permutation methods. NeuroImage. 2003 December; 20(4): p. 2343-2356, **doi: 10.1016/j.neuroimage.2003.08.003.**

[10] Hagler DJ, Saygin AP, Sereno MI. Smoothing and cluster thresholding for cortical surface-based group analysis of fMRI data. NeuroImage. 2006 December; 33(4): p. 1093-1103, **doi: 10.1016/j.neuroimage.2006.07.036.**

[11] Eklund A, Nichols TE, Knutsson H. Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. Proceedings of the National Academy of Sciences. 2016 July; 113(28): p. 7900-7905, **doi: 10.1073/pnas.1602413113.**

[12] Marroquin JL, Biscay RJ, Ruiz-Correa S, Alba A, Ramirez R, Armony JL. Morphology-based hypothesis testing in discrete random fields: A non-parametric method to address the multiple-comparison problem in neuroimaging. NeuroImage. 2011 February; 56(4): p. 1954-1967, **doi: 10.1016/j.neuroimage.2011.09.051.**

[13] Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers. 1st ed. Press P, editor. New York, USA.: Psychology Press; 2014, **doi: 10.4324/9781315806167.**

[14] Hochberg Y, Tamhane AC. Distribution-Free and Robust Procedures. 1st ed. Shube B, editor. New York, USA.: Wiley Online Library; 1987, **doi: 10.1002/9780470316672.**

[15] Pantazis D. General Linear Modeling of Magnetoencephalography Data. In Bronzino JD, Peterson DR, editors. Biomedical Signals, Imaging, and Informatics. Boca Raton, Florida, USA.: CRC Press; 2014. p. 33-350, **doi: 10.1201/b15468.**

[16] Benjamini Y. Discovering the false discovery rate. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2010 August; 72(4): p. 405-416, **doi:10.1111/j.1467-9868.2010.00746.x.**

[17] Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behavioural brain research. 2001 November; 125(1-2): p. 279-284, **doi: 10.1016/S0166-4328(01)00297-2.**

[18] Friston KJ, Ashburner JT, Kiebel SJ, Nichols TE, Penny WD. Statistical parametric mapping: the analysis of functional brain images. 1st ed. Penny W, editor. London, UK.: Academic press; 2011, **doi :10.1016/B978-0-12-372560-8.X5000-1.**

[19] Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping. 2002 January; 15(1): p. 1-25, **doi: 10.1002/hbm.1058.**

[20] Welvaert M, Rosseel Y. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fMRI data. PloS one. 2013 November; 8(11): p. 1-10, **doi: 10.1371/journal.pone.0077089.**

[21] Zhang H, Nichols TE, Johnson TD. Cluster mass inference via random field theory. Neuroimage. 2009 January; 44(1): p. 51-61, **doi: 10.1016/j.neuroimage.2008.08.017.**

[22] Rivera M, Ocegueda O, Marroquin JL. Entropy-controlled quadratic Markov measure field models for efficient image segmentation. IEEE Transactions on Image Processing. 2007 December; 16(12): p. 3047-3057, **doi: 10.1109/TIP.2007.909384.**

[23] Dalmau O, Rivera M. Beta-Measure for Probabilistic Segmentation. In Advances in Artificial Intelligence: 9th Mexican International Conference on Artificial Intelligence, MICAI 2010, Pachuca, Mexico, November 8-13, 2010, Proceedings, Part I; 2010; Berlin, Heidelberg: Springer Berlin Heidelberg. p. 312-324, **doi: 10.1007/978-3-642-16761-4_28.**

[24] Besag J. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society B. 1986 January; 48(3): p. 48-259, **doi: 10.1080/02664769300000059.**

[25] Geman S, Geman D. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1984 November; PAMI-6(6): p. 721-741, **doi: 10.1109/TPAMI.1984.4767596.**

[26] Boykov Y, Veksler O, Zabih R. Fast Approximate Energy Minimization via Graph Cuts. IEEE Trans. Pattern Anal. Mach. Intell. 2001 November; 23(11): p. 1222-1239, **doi: 10.1109/34.969114.**

[27] Wu YL, Agrawal D, Abbadi AE. Applying the golden rule of sampling for query estimation. ACM SIGMOD Record. 2001 May; 10(2): p. 449-460, **doi: 10.1145/376284.375724.**

[28] Devroye L. General Principles in Random Variate Generation. In Springer-Verlag , editor. Nonuniform random variate generation. New York, New York 10010, U.S.A: Elsevier; 2006. p. 27-36, **doi: 10.1016/S0927-0507(06)13004-2.**

[29] Drew JH, Evans DL, Glen AG, Leemis LM. Transformations of random variables. In Grassmann WK, editor. Computational Probability. Boston, MA, USA: Springer; 2017. p. 47-46, **doi: 10.1007/978-0-387-74676-0.**

[30] Oldham KB, Myland JC, Spanier J. The Error Function erf (x) and Its Complement erfc (x). In Science S, editor. An Atlas of Functions. New York, USA: Springer; 2008. p. 405-415, **doi: 10.1007/978-0-387-48807-3_41.**

[31] Besag J. Statistical Analysis of Non-Lattice Data. Journal of the Royal Statistical Society. Series D (The Statistician). 1975 January; 24(3): p. 179-195, **doi: 10.2307/2987782.**

[32] Friston K. Single subject epoch (block) auditory fMRI activation data [Internet]. 1999. Available from: *http://www. fil.ion.ucl.ac.uk/ spm/data/auditory.*

[33] Raz J, Zheng H, Ombao H, Turetsky B. Statistical tests for fMRI based on experimental randomization. NeuroImage. 2003 June; 19(2): p. 226-232, **doi: 10.1016/s1053-8119(03)00115-0.**

[34] Amunts K, Morosan P, Hilbig H, Zilles K. Chapter 36 - Auditory System. In Mai JK, Paxinos G, editors. The Human Nervous System (Third Edition). New York, USA.: Academic Press; 2012. p. 1270-1300, **doi: 10.1016/C2009-0-02721-4.**

[35] Bizley J. Audition. In Conn PM, editor. Conn's Translational Neuroscience. New York, USA: Elsevier; 2017. p. 579-598, **doi: 10.1016/B978-0-12-802381-5.00042-7.**