

Applications of random field theory to electrophysiology

James M. Kilner*, Stefan J. Kiebel, Karl J. Friston

The Wellcome Department of Imaging Neuroscience, Institute of Neurology, 12 Queen Square, London WC1N 3BG, UK

Received 6 July 2004; received in revised form 24 September 2004; accepted 20 October 2004

Abstract

The analysis of electrophysiological data often produces results that are continuous in one or more dimensions, e.g., time–frequency maps, peri-stimulus time histograms, and cross-correlation functions. Classical inferences made on the ensuing statistical maps must control family wise error (FWE) when searching across the map's dimensions. In this paper, we borrow multiple comparisons procedures, established in neuroimaging, and apply them to electrophysiological data. These procedures use random field theory (RFT) to adjust p -values from statistics that are functions of time and/or frequency. This RFT adjustment for continuous statistical processes plays the same role as a Bonferroni adjustment in the context of discrete statistical tests. Here, by analysing the time–frequency decompositions of single channel EEG data we show that RFT adjustments can be used in the analysis of electrophysiological data and illustrate the advantages of this method over existing approaches.

© 2004 Elsevier Ireland Ltd. All rights reserved.

Keywords: EEG; Wavelet; Oscillations; Electrophysiology

Time–frequency decomposition of EEG signals facilitates the study, in peri-stimulus time, of EEG components that occur in different frequency ranges [8,13,15]. Such analyses result in channel-wise 2-dimensional (2D) time–frequency maps. Commonly, researchers are interested in whether condition-specific effects, in a particular frequency range over some period of peri-stimulus time, are statistically significant. To test for these effects the individual time–frequency maps are often combined into a single statistical map (SPM), with one statistic per time–frequency bin [13]. However, inference on these 2D time–frequency statistical maps must account for the large number of multiple non-independent comparisons. In other words, significant effects must survive a threshold so there is only a small chance of false positives over all time–frequency bins. A simple, but inexact, method for setting this threshold is a Bonferroni correction. In practice, this method is rarely adopted because it assumes that all the statistics are independent. As a consequence, the corrected thresholds are too conservative for time–frequency SPMs where

there is a high degree of correlation among neighbouring bins.

In many analyses the multiple comparisons problem is circumvented by restricting the search space prior to inference such that there is only one statistic per repeated measure. This is usually accomplished by averaging the time–frequency data over a time–frequency window defined a priori [8,13,15]. A major limitation of this approach is that it requires the frequency range and time-period of the effect to be known in advance. A more natural approach would be to make inferences of the topography of the effect in time–frequency space using the SPM directly.

In the analysis of spatially correlated 3D fMRI and PET random field theory (RFT) is used to assign a p -value to topological features of the SPM such as its peak height [3,5,6,16–18]. This p -value is adjusted for the search window specified by the user. Recently, this approach has been adopted for SPMs of EEG [11] and MEG data [1]. RFT gives adjusted p -values by using results for the expected Euler characteristic (EC) of the excursion set of a smooth statistical process. The expected EC approximates the probability the map exceeds some height by chance. The ensuing p -values can be used to find a corrected height threshold (see [3,16]

* Corresponding author. Tel.: +44 020 7833 7457;
fax: +44 020 7813 1445.

E-mail address: j.kilner@fil.ion.ucl.ac.uk (J.M. Kilner).

for a more detailed introduction to RFT). One advantage of RFT is that it models continuous statistical processes and not a set of individual statistics. This means that RFT procedures can be used to characterise several topological features of the statistical map, e.g., maximal height, extent and number of clusters. The key intuition behind RFT procedures is that they control the false positive rate of maxima, or blobs, in statistical maps. A Bonferroni correction would control the false positive rate of bins, which would be inexact and unnecessarily conservative.

The assumptions under which the random field correction operates are quite simple and are, happily, satisfied by time–frequency data because of their inherent smoothness. RFT is a cornerstone of inference in human brain mapping that enables researchers to adjust their p -values to control false positive rates over spatial search volumes within the brain. As noted above, the key null distribution is that of the maximum statistic over the search volume. By evaluating any observed statistic, in relation to the null distribution of the maximum statistic, one is implicitly implementing a multiple comparisons procedure for continuous data. The distribution of the maximum statistic can be estimated non-parametrically using randomisation or permutation methods. More commonly, analytic forms of this distribution are derived using results from RFT. These results use the expected Euler characteristic of excursion sets above some specified threshold. For high thresholds this expectation is the same as the probability of getting a maximum statistic above threshold. By treating the data, under the null hypothesis, as continuous random fields the distribution of the Euler characteristic of any statistical process derived from these fields can then be used as an approximation to the null distribution required for inference. This assumes that the random variables (i.e., errors) conform to a good lattice approximation of an underlying random field. Furthermore, the expressions require that the random fields are multivariate Gaussian with a continuously differentiable autocorrelation function. It is a common misconception that this correlation function has to be Gaussian. It does not. Furthermore, the autocorrelation function does not have to be stationary. The ensuing p -value is a function of the search volume, over any arbitrary number of dimensions, and the smoothness (expressed in terms of full-width, half-maximum (FWHM)) of the underlying fields. A useful concept, that combines these two measures, is the resel or resolution element. The resel corresponds to the number of FWHM elements that comprise the search volume. Heuristically these correspond roughly to the number of independent observations. In other words, even a large search volume may contain a relatively small number of resels, if it is smooth. This calls for a much less severe adjustment to the p -value than would be obtained with a Bonferroni correction based on the number of bins, voxels or grid points. RFT now has an established role in imaging statistics. Some of the key papers in its development include [5,6,17]. Useful reviews of this will be found in the chapters by Brett et al. [3] and Worsley [16].

Here we use RFT, as implemented in SPM2 (Wellcome Department of Imaging Neuroscience, London, UK), to adjust p -values from SPMs of time–frequency data. The time–frequency maps were calculated from the EEG time-series from a single occipital electrode in two experimental conditions: one where the presentation of a visual stimulus was cued, C and a second, where it was not-cued, NC. Time–frequency maps were calculated for each trial, using a continuous wavelet decomposition (Morlet) from 5 to 70 Hz in a 750 ms time period prior to stimulus presentation. These time–frequency maps were averaged over trials and log transformed to produce one time–frequency of power per condition per subject. For each subject a contrast was taken as the difference between C and NC.

These contrasts were smoothed by convolution with a Gaussian kernel in both time (96 ms FWHM) and frequency (12 Hz FWHM) domains (see Fig. 1A for a schematic of the processing stream). This smoothing step is important for two reasons. First, it acts to blur any effects that are focal in the time and/or frequency dimensions, ensuring overlap among subjects. Secondly, it assures one of the assumptions of RFT: namely, that the errors conform to a good lattice approximation of a random field with a multivariate Gaussian distribution. The size of the Gaussian kernel used should reflect the heterogeneity of the locus of the peak effect in time–frequency space across subjects. Here the size of smoothing kernel was based on the reported variability over subjects in the latency and frequency of peak gamma-band activity [12,15].

In our example, the time–frequency SPM(T) for effects greater than zero was calculated from the smoothed contrasts using a one-sampled t -test (Fig. 1B). The peak time–frequency bin was 410 ms prior to the stimulus presentation at 42 Hz. However, although this peak bin had a Z -score of 3.36, which would be significant at $p < 0.0005$ uncorrected for family wise error (FWE), this bin did not reach the height threshold corrected for the entire time–frequency space ($p = 0.109$ corrected for FWE). Therefore, based on the height threshold, and in the absence of any a priori search window, we would not be able to infer any significant difference in modulation prior to the cued and non-cued stimuli. However, the RFT approach affords not only adjusted p -values based on height but also extent [6]. Extent-based inferences, based on the probability that a cluster of time–frequency bins could have occurred by chance can be more powerful than height based inferences, but at the expense of reduced localisation in time–frequency. For the SPM thresholded at $p < 0.01$ uncorrected, shown in Fig. 1B, there was a significant cluster of bins at $p < 0.05$ corrected (Fig. 1C). In this example, despite the absence of a priori information of where in time–frequency we should have expected any effect, the use of RFT, revealed a significant increase in oscillatory activity that occurred in the 33–62 Hz frequency range in a time window extending from 340 to 470 ms prior to stimulus presentation.

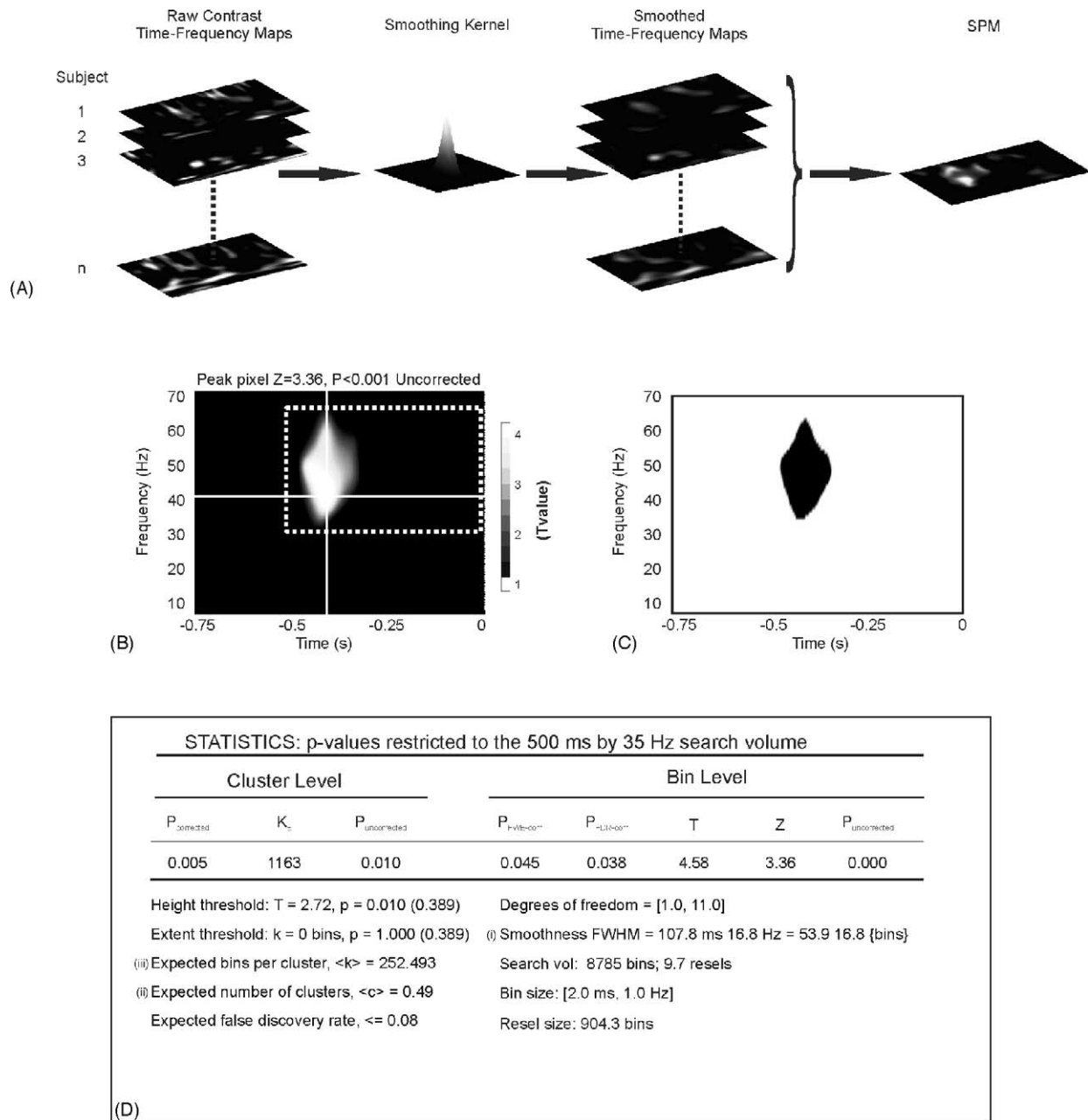
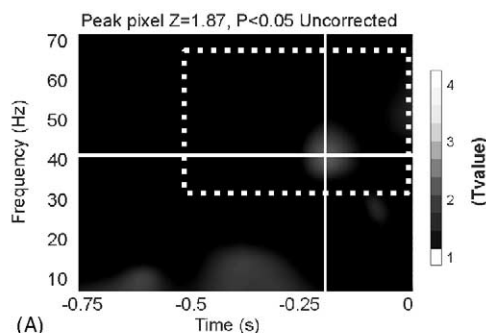


Fig. 1. (A) Shows the pipeline for the approach used here. Time–frequency contrasts were calculated for each subject and smoothed by convolution with a Gaussian kernel. These data were then analysed in SPM2 (Wellcome Department of Imaging Neuroscience, London, UK); (B) the SPM calculated from the smoothed time–frequency images and thresholded at $p < 0.01$ uncorrected. The location of the peak bin is shown. The white dotted box indicates our illustrative a priori window of interest; (C) the significant cluster of bins at $p < 0.01$ uncorrected; and (D) the results table from the analysis of the time–frequency SPM restricted to the window of interest as shown in B. With this analysis both the cluster level and bin level results are significant at $p < 0.05$ corrected for FWE. Note that although the data were smoothed with a Gaussian kernel of FWHM 96 ms and 12 Hz the smoothness calculated from the data was greater with FWHM of 107.8 ms and 16.8 Hz. (i) This difference reflects the correlation in the underlying data between adjacent bins in both the time and frequency dimensions. Note also that at a height threshold of $T = 2.72$ the expected number of clusters was 0.49, (ii) where one cluster was expected to be 252.493 bins, and (iii) whereas the cluster level analysis revealed one cluster of 1163 bins (K_E). Resel size means the number of bins in one resel. A resel is a “resolution element” and can be regarded as roughly equivalent to one independent observation.

In most instances searches over SPMs are constrained or directed. This is common in neuroimaging when we know where in the brain to look. For the example considered here we may want to constrain the search space to some expected time and frequency of the effect. In this instance the RFT

adjustment of the p -values would only be over bins within this search window. For example, had we defined a priori a window of interest 500 ms prior to stimulus presentation in the gamma-band range, 30–65 Hz (Fig. 1B and D) then the peak bin would have been significant at $p < 0.05$



STATISTICS: p-values restricted to the 500 ms by 35 Hz search volume

Cluster Level			Bin Level				
$P_{\text{corrected}}$	K_{E}	$P_{\text{uncorrected}}$	$P_{\text{FWE corr}}$	$P_{\text{LRT corr}}$	T	Z	$P_{\text{uncorrected}}$
0.613	523	0.650	0.533	0.968	2.07	1.87	0.031
Height threshold: $T = 1.36$, $p = 0.095$ (0.959)			Degrees of freedom = [1.0, 11.0]				
Extent threshold: $k = 0$ bins, $p = 1.000$ (0.959)			Smoothness FWHM = 158.6 ms 16.5 Hz = 79.3 16.5 (bins)				
Expected bins per cluster, $\langle k \rangle = 1215.822$			Search vol: 8785 bins; 6.7 resels				
Expected number of clusters, $\langle c \rangle = 1.46$			Bin size: [2.0 ms, 1.0 Hz]				
Expected false discovery rate, ≤ 0.97			Resel size: 1290.5 bins				

Fig. 2. (A) Shows the SPM calculated from the smoothed time–frequency images of a null contrast and thresholded at $p < 0.05$ uncorrected. The location of the peak bin is shown. The white dotted box indicates our illustrative a priori window of interest from Fig. 1B and (B) the results table from the analysis of the time–frequency SPM restricted to the within the window of interest shown in A. With this analysis neither the cluster level nor the bin level results are significant at $p < 0.05$ corrected for FWE.

corrected for FWE (details of this analysis are given in Fig. 1D).

To illustrate the importance of adjusting the p -value, we repeated the above analysis exactly but using a null contrast. This null contrast involved randomly exchanging or re-assigning the normalised time–frequency data to different levels of the experimental factor. The results of the null contrast are presented using the same format as the previous figure in Fig. 2 (note we have used a much lower threshold for the SPM). Inspection of the table (Fig. 2B) will show that the largest t value falling within the search volume (dotted white line) had an uncorrected p -value of 0.031. However, this is not a significant result because the adjusted p -value, corrected for FWE was 0.533. Incidentally, if we repeated the randomisation to produce a large number of null contrasts, and recorded the distribution of the maximum statistic within the search volume, we would have an empirical estimate of the null distribution required for inference. This is exactly how non-parametric versions of statistical parametric mapping are implemented.

This null example is presented in a purely heuristic way to highlight the importance of correcting for p -values. More thorough evaluations of SPM for log transformed power

data suggests that the approach is robust [10]. In brief, both smoothing, and the construction of contrasts, involve taking linear mixtures of (non-Gaussian) power data. By central limit theorem these mixtures are effectively Gaussian leading to valid and exact tests in Monte Carlo simulations [10].

Above we have commented upon inference based upon the peak height of the SPM and the spatial extent above a pre-specified threshold. The latter p -values were based on relatively simple expressions that assume the study has high degrees of freedom and that the smoothness is stationary over time–frequency space. This is easily motivated in the present context because the time–frequency data was smoothed with a large smoothing kernel prior to inference. However, in the absence of this smoothing better approximations are available (e.g., in FMRISTAT <http://www.math.mcgill.ca/keith/fmristat/>). More refined distributional approximations accommodate non-stationary smoothness [4] and uncertainty about the estimation of smoothness or FWHM (see the appendix of reference [7]).

In this introductory paper, we have focussed on multiple comparison procedures for continuous data based on RFT. As mentioned earlier, the requisite null distribution of the maximal statistic can be estimated using non-parametric proce-

dures. This is referred to as statistical non-parametric mapping. The analytic and closed form expressions provided by RFT are based on assumptions that render it more powerful or sensitive than equivalent non-parametric approaches. However, if these assumptions are violated non-parametric approaches can be employed. An interesting example of this is the application of statistical non-parametric mapping to source reconstructed MEG data at a particular time and frequency [14]. In this example, the reconstruction uses beamformer techniques that induce some badly behaved spatial correlations among remote sources that may confound RFT. For this reason the authors prefer non-parametric approaches. In the context of time–frequency analyses a well-behaved correlation structure is assured by the nature of the decomposition and by the post hoc smoothing described above.

This paper has focussed on the analysis of a single channel over time and frequency. The procedures described above complement the analysis of single time and/or time–frequency points that are continuous over spatially reconstructed electrical or magnetic sources [1,2,11,14]. In Kiebel and Friston [9] we consider high dimensional SPMs that encompass both space and peristimulus time to form a four-dimensional search space. Given that RFT can accommodate high dimensional search spaces it would, in principle, be possible to extend the time–frequency analyses described here to five dimensions (three spatial, time and frequency). However, in practical terms we anticipate that statistical parametric mapping will be applied over space, or over time–frequency but not both. There are both conceptual and computational reasons for this, which are detailed in [9]. This means that time–frequency analyses of the sort described above, would be performed on single channels of particular interest, the temporal expression of specified spatial modes of distributed activity or the electromagnetic responses at a particular point in the brain in source space.

In this paper, we have described the application of SPM to EEG data to allow statistical inferences in time–frequency space. The advantages of this approach are:

- (i) No a priori predictions about the location of effects are required to make statistically valid inferences.
- (ii) Statistical inferences are based on p -values adjusted for multiple comparisons.
- (iii) Statistical inferences can also be made on the extent of any effect, which may be more powerful than the height.
- (iv) The machinery for analysis and statistical inference are already fully established, widely used and available as academic software (<http://www.fil.ion.ucl.ac.uk/spm>).

Here we have focussed on time–frequency SPMs, however, this approach can also be applied to other multi-dimensional data such as peri-stimulus time histograms, time-dependent cross-correlations and time-dependent phase synchronisation maps. We will be using these inference devices in forthcoming publications.

Acknowledgement

The Wellcome Trust funded this work.

References

- [1] G. Barnes, A. Hillebrand, Statistical flattening of MEG beamformer images, *Hum. Brain Mapp.* 18 (2003) 1–12.
- [2] J. Bosch-Bayard, P. Valdes-Sosa, T. Virues-Alba, E. Aubert-Vazquez, E. John, T. Harmony, J. Riera-Diaz, N. Trujillo-Barreto, 3D statistical parametric mapping of EEG source spectra by means of variable resolution electromagnetic tomography (VARETA), *Clin. EEG Electroencephalogr.* 32 (2001) 47–61.
- [3] M. Brett, W. Penny, S. Kiebel, An introduction to random field theory, in: *Human Brain Function II*, Academic Press, London, UK, 2003 (Chapter 14).
- [4] J. Cao, K.J. Worsley, Applications of random fields in human brain mapping, in: M. Moore (Ed.) *Spatial Statistics: Methodological Aspects and Applications*, Springer Lecture Notes in Statistics, vol. 159, 2001, pp. 169–182.
- [5] K.J. Friston, C.D. Frith, P.F. Liddle, R.S.J. Frackowiak, Comparing functional (PET) images: the assessment of significant change, *J. Cereb. Blood Flow Metab.* 11 (1991) 690–699.
- [6] K.J. Friston, K.J. Worsley, R.S.J. Frackowiak, J.C. Mazziotta, A.C. Evans, Assessing the significance of focal activations using their spatial extent, *Hum. Brain Mapp.* 1 (1994) 214–220.
- [7] S. Hayasaka, K. Luan-Phan, I. Liberzon, K.J. Worsley, T.E. Nichols, Non-Stationary cluster-size inference with random field and permutation methods, *Neuroimage* 22 (2004) 676–687.
- [8] A. Keil, M.M. Muller, W.J. Ray, T. Gruber, T. Elbert, Human gamma band activity and perception of a gestalt, *J. Neurosci.* 19 (1999) 7152–7161.
- [9] S.J. Kiebel, K.J. Friston, Statistical parametric mapping for event-related potentials: I. Generic considerations, *Neuroimage* 22 (2004) 492–502.
- [10] S.J. Kiebel, C. Tallon-Baudry, K.J. Friston, Analysis of oscillatory activity as measured with eeg/meg using statistical parametric mapping, submitted for publication.
- [11] H. Park, J. Kwon, T. Youn, J. Pae, J. Kim, M. Kim, K. Ha, Statistical parametric mapping of LORETA using high density EEG and individual MRI: application to mismatch negativities in schizophrenia, *Hum. Brain Mapp.* 17 (2002) 168–178.
- [12] A. Posada, E. Hugues, P. Vianin, N. Franck, J. Kilner, Augmentation of induced visual gamma activity by increased task complexity, *Eur. J. Neurosci.* 18 (2003) 1–6.
- [13] E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, F.J. Varela, Perception's shadow: long-distance synchronization of human brain activity, *Nature* 397 (1999) 430–433.
- [14] K.D. Singh, G.R. Barnes, A. Hillebrand, Group imaging of task-related changes in cortical synchronisation using nonparametric permutation testing, *Neuroimage* 19 (2003) 1589–1601.
- [15] C. Tallon-Baudry, O. Bertrand, C. Delpuech, J. Pernier, Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans, *J. Neurosci.* 17 (1997) 722–734.
- [16] K.J. Worsley, Developments in random field theory, in: *Human Brain Function II*, Academic Press, London, UK, 2003 (Chapter 15).
- [17] K.J. Worsley, A.C. Evans, S. Marrett, P. Neelin, A three dimensional statistical analysis for CBF activation studies in the human brain, *J. Cereb. Blood Flow Metab.* 12 (1992) 900–918.
- [18] K.J. Worsley, S. Marrett, P. Neelin, A.C. Nandal, K.J. Friston, A.C. Evans, A unified statistical approach for determining significant voxels in images of cerebral activation, *Hum. Brain Mapp.* 4 (1996) 58–73.