Proceedings of the 2006 WSEAS International Conference on Mathematical Biology and Ecology, Miami, Florida, USA, January 18-20, 2006 (pp240-248)

New Approach to Nonparametric Statistical Analysis of fMRI Signals

PATRICK A. DE MAZIÈRE & MARC M. VAN HULLE
Laboratorium voor Pyscho- & Neurofysiologie
Medical School, Campus Gasthuisberg O & N
Katholieke Universiteit Leuven
Herestraat 49, bus 1021, 3000 Leuven
BELGIUM

Abstract: We present a method that enables the use of nonparametric EDF-like statistics for analysing fMRI data that is known to be autocorrelated over time. Analysis and comparison with existing methods like the common General Linear Model solution or a permutation test confirm its validity and usefulness. In addition, our method requires considerably less computation time than a permutation or Bayesian test.

Key-Words: fMRI, nonparametric statistics, permutation tests, GLM

1 Introduction

During the past few decades, neurosciences have enjoyed an ever-growing scientific interest. This could be explained by man's unrelenting urge to explore the unknown, but there is also the invention and availability of new and more powerful techniques and computers. The former have provided new scanners to record the brain *in vivo*, the latter made it possible to perform sophisticated statistical analyses at acceptable speeds. One of the scanning techniques most frequently used is *Magnetic Resonance Imaging* (MRI). The MRI scanner has the advantage that the anatomy and physiology can be visualised non-invasively, *i.e.*, without the use of noxious contrast agents or harmful radiations. The term *functional* MRI (fMRI) is employed whenever the functioning of organs is investigated by taking multiple successive scans over time.

The MRI signals are corrupted by a large number of noise sources, *e.g.*, physiological ones, mechanical ones, and those inherent to the fMRI principle. A consequence is that activation of a brain region causes only a 1-2% signal change with respect to the higher cognitive tasks using 3T scanners [1]. Therefore, the development of powerful analysis tools, which are able to cope with the specific fMRI signal properties, is still ongoing.

The tools used to analyse and explore an fMRI data set can be divided into two main categories: model-based and model-free ones. The best known representative of the model-based tools is the General Linear Model (GLM). A GLM represents every effect, which is assumed to be present in the recorded fMRI signal, by a single regressor which, in addition, is convolved with a so-called haemodynamic response function (HRF) to model the haemodynamic delay of the brain [2, 3]. An fMRI signal is thus represented by a linear combination of these regressors. The obtained GLM is solved using ordinary least squares (OLS); the obtained regression coefficients are then combined into a single statistical t-value, which expresses the responsiveness of the corresponding brain region with respect to the given stimuli. Ordinary Least Squares (OLS) is used to solve this GLM.

In this paper we focus on model-free methods and propose an extension for existing nonparametric statistics in order to make them applicable to fMRI data. Indeed, fMRI data has a significant serial correlation[a] that, when

[a] fMRI data is autocorrelated *both* over time and over space. To discern between both, the temporal autocorrelation is referred to as ~~*serial correlation*, while the~~ term *(spatial) autocorrelation* is reserved to indicate the spatial aspect.

it is not properly taken into account, prohibits a valid (nonparametric) statistical analysis. The novel approach we propose performs equally well in comparison to the well-known permutation test. Our method has a serious advantage over, *e.g.*, permutation test-based methods or Bayesian methods since it requires a negligible amount of computer time to apply (a couple of minutes versus several hours).

This article's structure is as follows: first, we discuss the reasons behind our choice for nonparametric statistics for analysing fMRI data sets. Second, we propose three nonparametric statistical tests: their definition, why serial correlations prohibit a valid analysis of fMRI signals using these nonparametric statistics, and how we solved this problem. Next, we discuss two methods to account for this serial correlation, and compare these methods to the results obtained with the permutation test. For the quantitative comparisons we use synthetic and real fMRI data sets. Finally, we conclude this article with a discussion and some general remarks.

2 Topics of Attention in Performing Statistical Analyses of fMRI Data

A topic of discussion concerns the question whether the recorded data can be validly analysed using Gaussian statistics or not. Since nonparametric tests are the only kind of statistical tests that are guaranteed to be valid and exact whenever the nature of the distribution is unknown [4], we explore the application of nonparametric statistical tests, and thus model-free analyses.

There is a fair amount of literature discussing the application of nonparametric statistical tests to neuroimaging data, mainly with respect to the significance assessment, and to a lesser extent for the detection of activation. As significance assessment one can choose for a permutation test [4, 5], the False Discovery Rate (FDR) [6, 7], or Bayesian techniques [8]. These methods are independent from the underlying distribution of the obtained statistical significance values, and are therefore applicable to the statistical values obtained with both parametric and nonparametric tests. As a nonparametric alternative for the detection of activation, the Kolmogorov-Smirnov test or a Wilcoxon-based variant[b] is often used [9, 10]).

However, the application of nonparametric tests to fMRI data is complicated by the presence of serial correlations. As reported by [9], the outcome of any (nonparametric) statistical test is invalid when these serial correlations are not taken into account. In case of the GLM-based test, the GLM's residue is analysed to obtain a correction factor that enables a veracious analysis [11, 12]. In case a nonparametric test is employed, no GLM is used, preventing us to extract a correction factor. Therefore, we introduce in this paper a novel method that allows us to take the (serial) autocorrelations into account and that is applicable to rank-order based nonparametric tests like the Kolmogorov-Smirnov test (*KS*), the Mann-Whitney two-sample test (*MW*), and the Cramér-von Mises two-sample test (*CvM*).

3 Nonparametric Test Statistics for Analysing fMRI Data

For the analysis of fMRI data, we selected rank tests that are based on an *empirical distribution function* (EDF). These rank tests are known to be the most powerful nonparametric ones [13]. Another argument to opt for this kind of tests becomes clear when we discuss the serial correlation problem. Amongst this category we count the *KS* and *CvM* statistical tests [13]. The *MW* test does not belong to this category, but the way its statistical values are obtained is very similar. Before we discuss each test in detail we explain how statistical information is extracted from fMRI data.

3.1 EDF Statistical Tests and fMRI (Multi-Condition) Experimental Designs

Traditionally, OLS solvers are applied to GLMs to estimate the regression coefficients. The regression coefficients, each representing the average activity level of a brain spot with respect to a given stimulus, are then combined with the regression error into one statistical value that expresses the relative responsivity of that brain spot. This procedure is fully described in, *e.g.*, [11].

With respect to nonparametric statistics where no regression coefficients are available, we compare the recorded values themselves. The here proposed method is

[b] The Mann-Whitney test does appear in literature under different names. Other frequently used names are the *Mann-Whitney U* test and the *Wilcoxon rank sum* test.

suitable for any nonparametric (rank-order) test that can test for a difference between two samples that possibly contain an unequal number of data points. To explain the method, assume a study using 6 different stimuli, labelled A through F, and where we want to test the responsiveness of a brain region in favour of stimuli A, B, and D over stimulus E (contrast = $A+B+D-E$). For this purpose we put in sample $\{X_i\}$ all data points recorded under stimuli A, B and D, while $\{Y_i\}$ contains those recorded under stimulus E. The fact that one sample can contain three times as many data points as the other is correctly dealt with by the respective statistical tests. A balancing of the samples is thus obsolete.

3.2 Mann-Whitney Two-Sample Test

This test checks for a difference in location, *i.c.*, the median. As such, it is the nonparametric counterpart of the parametric t test that checks for a difference in mean. Given, two samples of data points, $\{X_i\}$ and $\{Y_i\}$, containing N_x and N_y data points respectively, a set of data points $\{Z_i\} = \{X_i\} \bigcup \{Y_i\}$ is created and a rank assigned to the respective data points of $\{X_i\}$ and $\{Y_i\}$. The statistical value T_1 is obtained as shown in (2), where $N = N_x + N_y$ and $\sum_{i=1}^{N} R_i^2$ represents the sum of squares of *all* N ranks. The significance values p are easily calculated since T_1 is approximately a standard normal random variable [13].

$$\begin{aligned} T &= \sum_{k=1}^{N_x} R(X_k) && (1) \\ T_1 &= \frac{T - N_x \frac{N+1}{2}}{\sqrt{\frac{N_x N_y}{N(N-1)} \sum_{i=1}^{N} R_i^2 - \frac{N_x N_y (N+1)^2}{4(N-1)}}}. && (2) \end{aligned}$$

3.3 Kolmogorov-Smirnov & Cramér-von Mises Two-Sample Test

Using the same notation, $\{X_i\}$ and $\{Y_i\}$ are represented by their EDF: $S_1(x)$ and $S_2(x)$, respectively. The EDF $S(x)$ represents the fraction of X_is that are less than or equal to x [13]. An EDF statistical test then verifies the hypothesis that both samples are drawn from the same distribution, based on the deviations between the EDFs: $d_k = S_1(x_k) - S_2(x_k)$, for $k = 1, \ldots, (N_x + N_y)$.

$$T_2 = \frac{N_x N_y}{(N_x + N_y)^2} \sum_{x_k \in \{X_i\} \cup \{Y_i\}} \Big(S_1(x_k) - S_2(x_k) \Big)^2 \qquad (3)$$

The difference between the *CvM* and *KS* test is that the calculation of the statistical values is based on all d_ks for the *CvM* test (3), while it is simply $\sup(|d_k|)$ for the *KS* test. The difference in definition causes also a difference in the range of the statistical values: $[0,1]$ for the *KS* test, and $[0,\infty)$ for the *CvM* test[c].

Three assumptions must be satisfied when applying the *KS*/*CvM* and *MW* tests [13]: the measurement scale should be ordinal, the random variables should be continuous and the data points should be exchangeable. Only, the third assumption does not hold for fMRI time series since serial correlations are present. Consequently, the significance thresholds calculated theoretically for those EDF statistical tests are not valid since they are derived under the assumption of white noise. In the next section we first discuss the concept of serial correlations, where after we propose our solution to correctly analyse correlated data using rank order tests.

3.4 Rank-Order Statistical Tests and Serial Correlation

3.4.1 Serial Correlations in fMRI

Serial correlations are characterised by two parameters: the *lag* τ, and the amount of (auto) correlation ρ per lag as shown in (4) for a first order autoregressive model (AR(1)). The lag parameter τ expresses the time over which the value of x at time t is influenced by another one, while ρ expresses the amount of influence. In general, the actual value of a fMRI signal $x(t)$ is measured by a number of values from the past as formulated in (5) (σ represents the standard deviation):

$$\begin{aligned} x(t) &= \rho_1 x(t-1) + u(t), \qquad 0 \leqslant \rho_i \leqslant 1 && (4) \\ x(t) &= \rho_1 x(t-1) + \ldots + \rho_\tau x(t-\tau) + u(t) && (5) \\ &\quad \text{where } u(t) \text{ is normally distributed with} \\ &\qquad \bar{u}(t) = 0, \quad \sigma^2_{u(t)} = \text{constant } \forall t \\ &\qquad \sigma_{u(t)u(t-s)} = 0 \qquad \forall t, \forall s \neq 0 \end{aligned}$$

Many of the current (GLM-based) fMRI analysis tools, adopt a two-stage pre-whitening that corrects for serial

[c] The calculation of the significance value $p(T_2)$, *i.e.*, the probability that two empirical distributions are drawn from the same population distribution, is rather complex and can be found in [14, 15] or received upon request from the author.

correlations, to fulfill the requirement that the residuals of the GLM must be independent and identical distributed (iid) [16]. This pre-whitening procedure first estimates the autocorrelation exploring the residuals of an initial model fit. In a second step the estimated autocorrelation is then removed from both the fMRI signal and the model. A well-known pre-whitening method is the Cochrane-Orcutt method [17, 16]. We further refer to this serial correlation corrected method as the OLS-CO method.

3.4.2 Methods for Correctly Applying Nonparametric Tests

Contrary to the GLM that relies on the OLS(-CO), nonparametric statistical tests do not rely on a GLM and cannot use therefore the obtained residuals to correct for the presence of serial correlations. Therefore, we developed a completely new method, at least to our knowledge, which enables us to correct for serial autocorrelations in case EDF-like statistical tests are used. It is based on the value τ_{max} that represents the maximum lag one wants to correct for. From literature, it is known that fMRI signals on average do not exceed an amount of autocorrelation $\rho = 0.4$ at lag one. The autocorrelation at higher lags is rather negligible, although some authors do correct for it as well [12] using AR models of second or higher order. The technique we propose here is theoretically applicable for any value of τ_{max}.

Without loss of generality, we assume here that only a lag one autocorrelation correction is necessary, and that the contrast equals $A - B$. According to the method presented in section 3.1, $\{X_i\}$ contains those data points that are recorded during stimulus A, and $\{Y_i\}$ those recorded during stimulus B. We limit this discussion here to the sample $\{X_i\}$ since $\{Y_i\}$ can be treated analogously. We divide $\{X_i\}$ into $\tau_{\max} + 1 = 2$ parts, labelled $\{X_i^{1*}\}$ and $\{X_i^{2*}\}$ according to the formula:

$$\begin{aligned} X_i^{1*} &= X_{2k} \\ X_i^{2*} &= X_{2k+1} \end{aligned} \tag{6}$$

with[d] $k = 1, \dots, \lfloor N_X/(\tau_{\max}+1) \rfloor$. This separation makes that the data points in $\{X_i^{1*}\}$ (or $\{X_i^{2*}\}$) have no longer the original lag one correlation and are thus exchangeable with each other. The p-values are then calculated for both $\{X_i^{1*}\}$ and $\{X_i^{2*}\}$ using either the *MW*, *KS*, or *CvM* test.

We now need a method that combines the $(\tau_{\max} + 1)$ p-values into a single p-value for the complete fMRI time signal. Two methods are discussed here. Given the fact that two p-values are calculated, a simple multiple comparison correction (MCC) can be used. Another approach is quite often used in the field of experimental psychology and is called *meta-analysis*. Meta-analyses allow combining two or more results obtained from possibly different groups to obtain an increased level of power. In fact even p-values obtained using different statistical tests can be used, as long as the hypothesis tested for is identical. Before we discuss the results obtained with each method, we explain both methods theoretically in the next paragraphs.

Multiple Comparison Correction Method to Combine p-Values Given the idea that the different p-values are obtained by performing identical statistical tests, the choice for using a multiple comparison correction (MCC) is rather obvious. Personal communication with Benjamini and Yekutieli, the authors of the already mentioned FDR, confirmed that a simple ”Simes FDR test for the intersection hypothesis” is valid to obtain a single corrected p^*-value. Such Simes test can be described as follows:

1. Order the p-values as follows: $p_1 \leq p_2 \leq \dots \leq p_{(\tau_{max}+1)}$.
2. $\forall i, \exists j \mid p^* = p_j \times (\tau_{max}+1)/j$ *and* $p_j \geq \max_i (p_i \times (\frac{j}{i}))$.

The obvious disadvantage of this technique is a decrease in sensitivity as we will see in the next sections where we display the outcome of some experiments.

Meta-Analysis to Combine p-Values Meta-analysis is described as the analysis of analyses [18, 19]. It is the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings. We use here the Stouffer combined test, which is very easy to interpret and to implement. With respect to our case, the question arises whether we can consider the different p-values, which we extracted from the different partial time series, as exchangeable. We will verify this in paragraph 3.6.

[d] $\lfloor x \rfloor$ or $floor(x)$ gives the largest integer $\leqslant x$.

The Stouffer combined test [19] converts the p-values into z-values. Indeed, given the fact that every p-value has an identical probability to occur, p-values are uniformly distributed and can therefore be transformed into z-values under the null hypothesis. The obtained z-values are summed properly (7) and transformed into a single p-value. We denote the z-value derived from the p_i-value of every partial time series by z_i, with $i = 1, \ldots, (\tau_{max} + 1)$. The global z-value, denoted by Z_c, for the complete fMRI time series can then be calculated as:

$$Z_c = \sum_{i=1}^{\tau_{max}+1} \frac{z_i}{\sqrt{\tau_{max}+1}}, \tag{7}$$

where $\tau_{max} + 1$ equals the number of tests combined, thus the number of partial time series examined. This procedure is based on the sum of the normal deviates being itself a normal deviate, with the variance equal to the number of observations summed [19]. The global p-value, P_c, can be derived very easily from Z_c.

3.5 Material and Methods for Validation

We start this paragraph with a discussion of the used data sets, where after we briefly discuss the permutation test: this test, together with a GLM-based test, is used to compare our results with.

Two kinds of data sets are used: a synthetic one, which is constructed using Gaussian noise that is autocorrelated using an AR(1) model that very well resembles the autocorrelation structure of real fMRI data ($\rho_1 = 0.4$), and a hybrid one that is constructed using fMRI null data, which consists of fMRI signals that are recorded while the volunteer in the scanner was at rest and not subject to any stimulus. FMRI-like signals are then obtained by adding a synthetic block pulse to both kinds of noise signals.

Synthetic data sets have the advantage that their properties are exactly known, while hybrid data sets better correspond with the real world situation. The hybrid data set we use here is extracted from the fMRI images made publicly available by the Brain Mapping Unit (University of Cambridge, UK). In order to extract noise signals from these images, we first pre-processed them: realignment using SPM99 software (Statistical Parameter Mapping, London, UK), followed by a grey matter segmentation using the FSL brain extraction tool (FMRIB Software Library, Oxford, UK). Noise signals are then generated by randomly extracting time signals from the grey matter images. Finally, a second order polynomial detrending and unit standardisation are applied to each one of them. The last two steps are also applied to the synthetic data sets.

We used both data sets to examine both the false positives rate (FPR) and the sensitivity or the true positives rate (TPR). For the FPR scenario, a bare noise signal is used to which *no* block pulse (see further) is added but which is examined as if a block pulse is present. Therefore, the null hypothesis states that no activation is present, and a rejection of this null hypothesis refers to a *false* positive. For the TPR scenario, an on-off block pulse is used to which a noise signal is added and that is examined as such. The null hypothesis remains identical, but now, a rejection of this null hypothesis refers to a *true* positive. The block pulse, which is used in the FPR & TPR testing scenarios, is an on-off block pulse train with 14 blocks of 30 scans each and with a repetition time between successive measurements equal to TR = 3s. To better mimic real fMRI signals, we convolved the bare block-pulse train with an HRF (HD = 7s) [2].

With respect to the nonparametric statistical tests where no GLM is used to analyse the signals, we have opted for a very simple approach to cope with this HRF: we left the transitional scans out from the analysis, *i.e.* the first $\lceil \mathrm{HD}/\mathrm{TR} \rceil$ data points[e] of every block are skipped. Experiments have shown that leaving out these transitional scans increases the power and performance of the statistical tests (results not shown).

Besides the OLS-CO test, we compare our novel approach also with the permutation test. The permutation test used here is the one introduced for fMRI by [4]. It allows to express the statistical significance using the data itself as a null distribution and is therefore a better point of reference than the OLS/t-test that is known to be too optimistic [20]. The test permutes the labels (conditions) rather than the measurements themselves. This guarantees that the serial correlation structure is preserved within each permuted time series. To obtain a reliable null distribution we opted to draw $1,000$ permutations for each fMRI signal. We calculate for every permuted time series the statistical value, being it either the OLS/t value, the T_1, KS, or the T_2 value. The statistical significance value (p) is then defined as the ratio of statistical values smaller

[e] $\lceil x \rceil$ or $ceil(x)$ gives the smallest integer $\geqslant x$, with $x \in \mathbb{R}$.

than the one of the original time series.

3.6 Validation and Comparison

To obtain reliable results, we based every statistical value (or values and conclusions thereof derived) on $10,000$ time series or iterations. This allows us to use the statistically common threshold of 0.01, since this nominal α guarantees that, at least theoretically, 100 cases should pass the test which is a significant amount to be detected properly. First, we discuss the TPR & FPR results obtained for the synthetic data set, next we discuss both curves for the hybrid data set.

3.6.1 Results for the Synthetic Data Set

For the sake of clarity, we first present in Fig. 1 the TPR curves for all statistical tests, and their corresponding permutation tests. At first glance, Fig. 1 indicates that the *MW*- and *CvM*-tests have more power than the *KS*-test. Since the *MW* turned out to be the most powerful, and to avoid unclear figures we display in the subsequent figures the *MW* test as only nonparametric test. Based on our simulations we can also state that analog conclusions are valid for the *CvM* and *KS* tests. We included the *CvM* test in our examination since we noticed overall that the *CvM* test outperforms the *KS* test (Fig. 2).

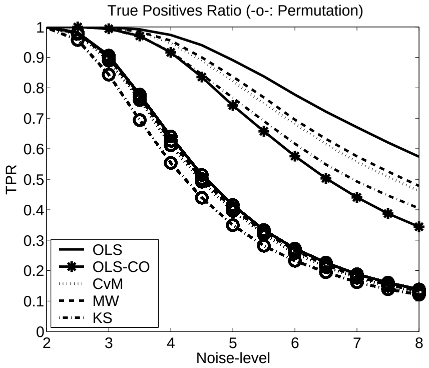

Figure 1: TPR curves for the OLS/*t*, *MW*-, *KS*-, and *CvM*-test and their corresponding permutation tests as a function of the amount of noise (noise-level, *x*-axis) using synthetic data. The curves for the permutation tests have a circular marker and a line-style identical to the line-style of the statistical test on which the permutation test is based. A value of one on the *y*-axis corresponds to 100% true activations.

If we combine the TPR results with the FPR values for the uncorrected case ($x = 0$ case at Fig. 2(d)), we must conclude that the OLS/*t* test fails the nominal size of rejections (0.01) as was to be expected. The OLS-CO/*t* test already better controls the FPR than the OLS/*t* test. The FPR values for the uncorrected nonparametric statistical tests deviate in a severe way from the nominal size. This is in agreement with the statement that the number of false positives (for the *KS*-test) is higher than that of the *t*-test [9]. Our simulations confirm this and extend this finding to the *MW* and *CvM* (not shown) tests.

To better control the FPR for the nonparametric tests we now examine the results obtained using either of the correction schemes: the first column of Fig. 2 displays the results obtained with the FDR scheme, while the second column shows the results obtained with the meta-analysis scheme. We restricted these figures to the curves for the OLS/*t*, OLS-CO/*t*, and *MW* statistical tests to avoid unclear figures[f]. We notice that the FPR is clearly decreased independent of the correction scheme used. Comparing both schemes with each other, we see that only the FDR scheme guarantees that the nominal size, 0.01, is achieved. The meta-analysis scheme clearly fails to control the FPR correctly for reasonable values of the lag parameter. Other simulations (not shown here) confirm this conclusion also for the *KS*- and *CvM*-test, and also show that, with respect to both the FPR and TPR values, the *MW* and *CvM* perform almost similar.

Considering again the FDR correction scheme and a correction up to lag 3, the power of the nonparametric statistical *MW* test approaches that of the permutation tests. As shown in Fig. 2(a), the TPR curves for the nonparametric and permutation tests almost coincide for this correction. Considering the FPR values with respect to this synthetic data set and the FDR correction scheme, Fig. 2(d) shows that also the FPR values nearly coincide, while that of the OLS-CO/*t* is slightly larger than the nominal size. Comparing the nonparametric statistical tests with the OLS-CO/*t* test we can summarise that a lag 1 correction has already a better false positive control than the standard OLS/*t*-test, that a lag 2 correction suffices to obtain a false positive rate equal to that of the OLS-CO/*t*-test, but that only a lag 3 correction returns a reasonable false positives control.

[f] Figures for the *KS* and *CvM* statistical tests, can be received upon request from the author.

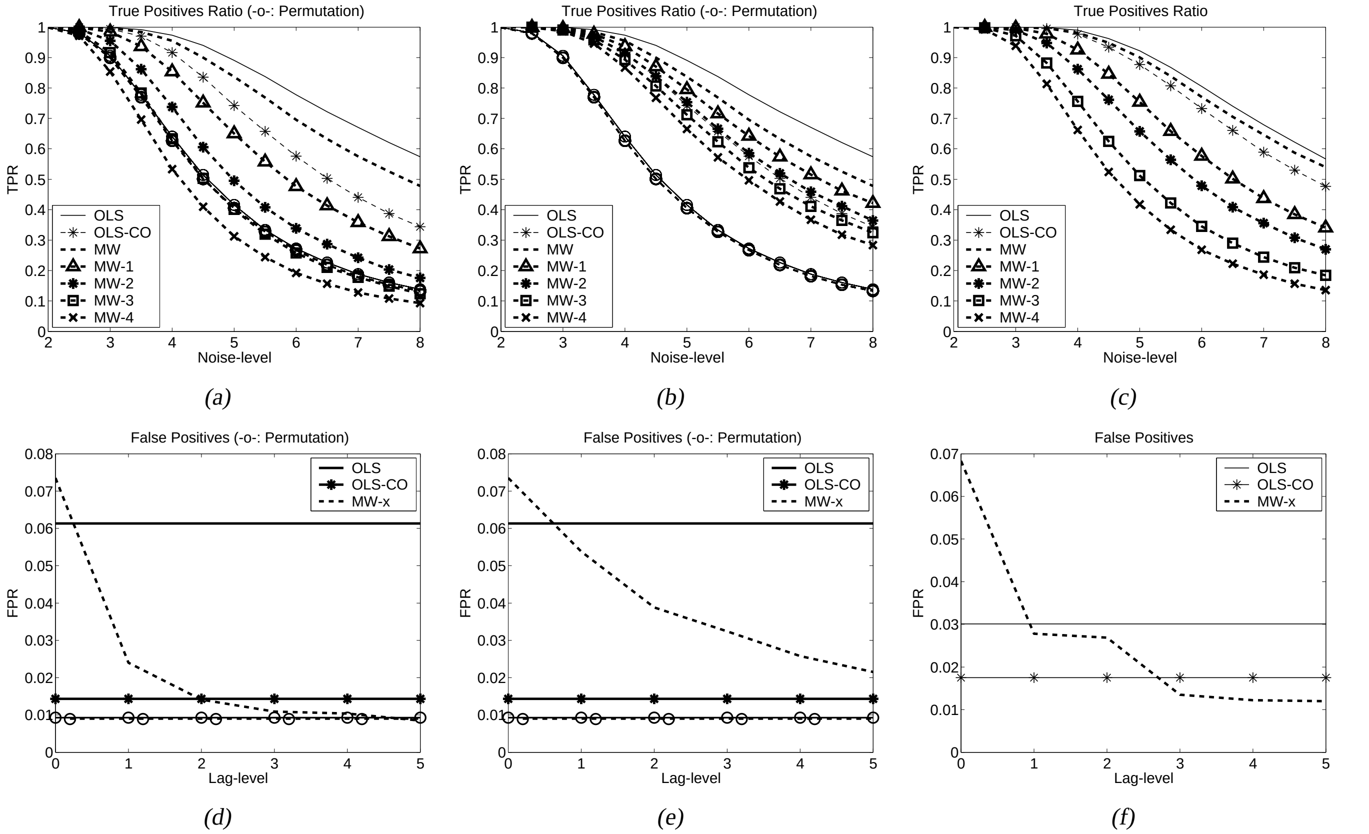

Figure 2: TPR & FPR curves for the OLS(-CO)/t, *MW* tests. The first two columns are the results obtained using the AR(1) synthetic data set, the third column (*(c)* & *(f)*) contains the results obtained for the hybrid data set. Figs. *(a)* to *(c)* display TPR curves as a function of the amount of noise (x-axis represents the amount of noise added to the block pulse, expressed in units standard deviation), *(d)* to *(f)* display the FPR curves (y-axis, $1 \equiv 100\%$) for corrections in the discrete range $\tau_{max} = [1, 5]$ (x-axis). The figures in the first and third column are obtained with the FDR scheme, while those of the second column are obtained with the meta analysis correction scheme. The lag we have corrected for is represented by the value behind the *MW* notation in the legend. If no value is given, no correction is applied.

3.6.2 Results for the Hybrid Data Set

Using synthetic data sets the properties of which are well known, we found that only the FDR scheme with a lag 3 correction seems to have a good FPR control. We now repeat the same validation/comparison procedure with respect to the hybrid data set to verify whether the same scheme and correction level still offer a good FPR control. These results are shown in the third column of Fig. 2.

We can deduce the following items from these and the previous figures: first, and in correspondence to what theory predicts [13], we see that when the data is derived from a Gaussian process (Fig. 1), the classic OLS test outperforms the nonparametric tests with respect to its power. Contrary, if the data is derived from a non-Gaussian process, as is the case with the hybrid data sets, the nonparametric statistical test better matches the OLS(-CO)/t TPR values. Second, when we compare the TPR & FPR curves, we see again that a lag 3 FDR correction has a better FPR control than the OLS-Co/t test.

By investigating the autocorrelation coefficient plot, one can define the value of τ_{max} as the lag value for which the autocorrelation coefficient drops below a given threshold. For both the synthetic AR(1) autocorrelated Gaussian noise and this hybrid data set, a correction at lag 3 seems to control the false positive rate rather well.

3.7 Discussion and Conclusion

Traditionally, a General Linear Model is used to analyse fMRI data. However, the question arises whether such analysis is valid given the Gaussian and linear assumptions underlying these methods. Specific pre-processing operations like data-smoothing [11] can help the data meet the required assumptions. We started investigating nonparametric statistical tests, to circumvent the question itself; moreover, nonparametric tests are the only kind of statistical tests that are guaranteed to be valid and exact in case the nature of the distribution is unknown [4].

Permutation tests and Bayesian techniques are good alternatives but require a huge amount of computing time. For this reason we examined whether classic nonparametric statistical tests can be adapted for application to fMRI data, which is known to be serial autocorrelated. We focused in this article especially on the Mann-Whitney test, although our research confirms that the analog conclusions hold for the Kolmogorov-Smirnov and the Cramér-von Mises test. To cope with the temporal autocorrelations, we developed and examined two possible methods to control the FPR of these nonparametric tests: one based on the False Discovery Rate (FDR, a multiple comparison correction method), the other based on meta-analyses. Using realistic synthetic data sets, we investigated both. The meta-analysis scheme, albeit promising when considering the TPR values (Fig. 2(b)), clearly fails to control the false positive rate sufficiently. Consequently, only the FDR correction scheme fulfills our needs. Since the time needed to perform this serial correlation correction method is within the order of seconds, our method has a clear advantage with respect to, *e.g.*, the permutation test or Bayesian techniques.

Furthermore, a current path of research might render the FDR technique even more promising: Yekutieli & Benjamini (personal communication, [21]) are developing hierarchical extensions to the basic FDR principle which allows to include information gathered while investigating part of the problem (*i.c.*, the serial correlation correction method), into the procedure that calculates adjusted p-values for the complete problem (multiple comparison correction). This path of research might return a solution yielding a higher sensitivity while keeping the FPR still within bounds.

With respect to the lower sensitivity of the nonparametric tests (even the permutation test) in comparison to the OLS-CO/t approach, we can mention two items in defence of them: 1) using an identical HRF model for both the creation of the synthetic signals and their analysis, which is in practice never the case, favours the OLS method, 2) we remind the reader that the OLS-based tests are rather optimistic [20]. This optimistic behaviour is confirmed by the OLS/t permutation test that has clearly less power than the OLS(-CO)/t-test.

Another minor disadvantage of our method is the fact that it is only applicable for block design fMRI studies and not for event-related fMRI studies. In addition, and contrary to the GLM based methods, nonparametric statistical tests do not allow to model additional effects such as eye movements or cardio-respiratory movements. A possible solution with respect to the nonparametric tests is to apply a statistical test that checks for any relationship between the selected time series and any of the effects. A warning for the researcher can then be issued in case a given threshold is surpassed.

Last, our approach allows also the use of statistical tests that reveal additional information about the detected activation. [10] mentioned already that application of a range of statistical procedures, parametric and data-driven, linear and nonlinear, would be most useful. Regions might show an equal average level of activity, but a different distribution of the observed activation. The application of *e.g.*, the Cramér-von Mises in addition to a Mann-Whitney test is therefore certainly a source of additional information for the researcher.

In conclusion, we can state that we have developed a method that enables the application of EDF-like nonparametric tests to fMRI data by accounting for the presence of serial correlations. Our method also enables the use of statistical tests that check for a difference in distribution and that thus return additional information to the researcher. In addition, certain pre-processing steps like data-smoothing, which tamper the data significantly, can be omitted. Finally, our method requires considerably less computation time (order of seconds) with respect to other nonparametric tests like the permutation or Bayesian tests.

Acknowledgement P.A. De Mazière is currently supported by a scholarship from the European Commission (NEST-2003-012963). M.M. Van Hulle is supported by research grants received from the Belgian Fund for Scientific Research – Flanders (G.0248.03 and G.0234.04), the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P5/04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), and the European Commission (NEST-2003-012963 and STREP-2002-016276).

References

[1] Jezzard P. *Physiological Noise: Strategies for Correction*, chap. 16. Springer, New York, 1999, pp. 173–181.

[2] Aguirre G., Zarahn E., and D'Esposito M. The Variability of Human, BOLD Hemodynamic Responses. *Neuroimage*, vol. 8(4), 1998, pp. 360–369.

[3] Villringer A. *Physiological Changes During Brain Activation*, chap. 1. Springer, New York, 1999, pp. 1–13.

[4] Holmes A., Blair R., Watson J., and Ford I. Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *J Cereb Blood Flow Metab*, vol. 16(1), 1996, pp. 7–22.

[5] Nichols T. and Holmes A. Nonparametric Permutation Tests For Functional Neuroimaging: A Primer With Examples. *Hum Brain Mapp*, vol. 15(1), 2001, pp. 1–25.

[6] Benjamini Y. and Hochberg Y. Controlling the False Discovery rate: a Practical and Powerful Approach to Multiple Testing. *Journal of Royal Stat. Soc. B*, vol. 57(1), 1995, pp. 289–300.

[7] Genovese C., Lazar N., and Nichols T. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *Neuroimage*, vol. 15(4), 2002, pp. 870–878.

[8] Woolrich M., Behrens T., Jenkinson M., and Smith S. Multi-level Linear Modelling for fMRI Group Analysis using Bayesian Inference. *Neuroimage*, vol. 21, 2004, pp. 1732–1747.

[9] Aguirre G., Zarahn E., and D'Esposito M. A Critique of the Use of the Kolmogorov-Smirnov (KS) Statistic for the Analysis of BOLD fMRI Data. *Magn Reson Med*, vol. 39(3), 1998, pp. 500–505.

[10] Lange N., Strother S., Anderson J., Nielsen F., Holmes A., Kolenda T., Savoy R., and Hansen L. Plurality and Resemblance in fMRI Data Analysis. *Neuroimage*, vol. 10(3), 1999, pp. 282–303.

[11] Worsley K., Liao C., Aston J., Petre V., Duncan G., Morales F., and Evans A. A General Statistical Analysis for fMRI Data. *Neuroimage*, vol. 15(1), 2002, pp. 1–15.

[12] Woolrich M., Ripley B., Brady M., and Smith S. Temporal Autocorrelation in Univariate Linear Modeling of fMRI Data. *Neuroimage*, vol. 14(6), 2001, pp. 1370–1386.

[13] Conover W. *Practical Nonparametric Statistics*. Third ed. John Wiley & Sons, New York, 1999. ISBN 0-471-16068-7.

[14] Anderson T. and Darling D. Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes. *Annals of Mathematical Statistics*, vol. 23, 1952, pp. 193–212.

[15] Csörgo S. and Faraway J. The Exact and Asymptotic Distributions of Cramér-von Mises Statistics. *J R Stat Soc [Ser B]*, vol. 58(1), 1996, pp. 221–234.

[16] Thomas R. *Modern Econometrics, an introduction*. Addison-Wesley, Harlow, UK, 1997.

[17] Cochrane D. and Orcutt G. Application Of Least Squares Regression To Relationships Containing Autocorrelated Error Terms. *Journal of the American Statistical Association*, vol. 44, 1949, pp. 32–61.

[18] Glass G. Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, vol. 5, 1976, pp. 3–8.

[19] Wolf F. *Meta-Analysis: Quantitative Methods for Research Synthesis*. Quantitative Applications in the Social Sciences. Sara Miller McCune, Sage Publications, Inc., Newbury Park, 1990.

[20] Gautama T. and Van Hulle M. Optimal Spatial Regularisation of Autocorrelation Estimates in fMRI Analysis. *Neuroimage*, vol. 23, 2004, pp. 1203–1216.

[21] Benjamini Y. and Yekutieli D. Hierarchical FDR Testing of Trees of Hypotheses. *Tech. Rep. 02-02*, Tel Aviv University, Department of Statistics and OR, 2003.