NeuroImage

www.elsevier.com/locate/ynimg
NeuroImage 26 (2005) 635–641

Technical Note

Spatial smoothing of autocorrelations to control the degrees of freedom in fMRI analysis

K.J. Worsley*

Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montreal, Québec, Canada H3A 2K6
McConnell Brain Imaging Centre, Montreal Neurological Institute, 3801 University Street, Montreal, Québec, Canada H3A 2B4

Received 16 September 2004; revised 3 January 2005; accepted 1 February 2005
Available online 24 March 2005

In the statistical analysis of fMRI data, the parameter of primary interest is the effect of a contrast; of secondary interest is its standard error, and of tertiary interest is the standard error of this standard error, or equivalently, the degrees of freedom (*df*). In a ReML (Restricted Maximum Likelihood) analysis, we show how spatial smoothing of temporal autocorrelations increases the effective *df* (but not the smoothness of primary or secondary parameter estimates), so that the amount of smoothing can be chosen in advance to achieve a target *df*, typically 100. This has already been done at the second level of a hierarchical analysis by smoothing the ratio of random to fixed effects variances (Worsley, K.J., Liao, C., Aston, J.A.D., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage, 15:1–15); we now show how to do it at the first level, by smoothing autocorrelation parameters. The proposed method is extremely fast and it does not require any image processing. It can be used in conjunction with other regularization methods (Gautama, T., Van Hulle, M.M., in press. Optimal spatial regularisation of autocorrelation estimates in fMRI analysis. NeuroImage.) to avoid unnecessary smoothing beyond 100 *df*. Our results on a typical 6-min, TR = 3, 1.5-T fMRI data set show that 8.5-mm smoothing is needed to achieve 100 *df*, and this results in roughly a doubling of detected activations.

Keywords: fMRI; Spatial smoothing; Autocorrelations; Degrees of freedom

Introduction

One of the simplest models for the first level of fMRI data analysis is the linear model with AR(p) temporal correlation structure (Bullmore et al., 1996; Locascio et al., 1997; Marchini and Smith, 2003; Woolrich et al., 2001; Worsley et al., 2002). The AR(p) parameters are estimated separately at each voxel, then spatially smoothed to reduce their variability, at the cost of slightly increasing their bias (see Fig. 1). Until recently, the amount of smoothing was chosen heuristically (e.g., 15 mm by the FMRISTAT software), but Gautama and Van Hulle (2004) have now introduced a method to estimate the amount of smoothing in a principled fashion. Their method is based on choosing the amount of smoothing to best predict the autocorrelations.

In this paper, we show how the amount of smoothing influences the effective degrees of freedom (*df*) of ReML (Restricted Maximum Likelihood) estimators (Harville, 1974). The theory uses the same techniques as Kenward and Roger (1997) and Kiebel et al. (2003). Kenward and Roger were concerned with ReML estimators (close to what FMRISTAT and FSL use), whereas Kiebel applied these techniques to least-squares estimators (close to what SPM uses). We show how the *df* can be approximated *before* the analysis is carried out, so that in conjunction with other regularization methods (e.g., Gautama and Van Hulle, 2004), we can choose the amount of smoothing in advance to achieve a targeted *df*. Curiously enough, this *df* is different for different contrasts, and different again for an F statistic that combines contrasts, so if more than one inference is desired from a single smoothing, then we suggest being conservative and taking the maximum. The proposed method is extremely fast and, unlike the method of Gautama and Van Hulle (2004), it does not require any image processing.

Method

We adopt a linear model for the mean of the fMRI data, with an AR(p) model for the temporal correlation structure (Eq. (6)). This is fitted by the pre-whitening method used by Worsley et al. (2002), similar to that used by the FSL software, and close to ReML, which is summarized as follows. The model is first fitted by least-squares to find least squares residuals. These residuals are used to calculate the temporal autocorrelations, and a simple method is used to correct for the bias incurred by using residuals from the fitted rather than the true model. The temporal autocorrelations are spatially smoothed to reduce their variability, then inserted into the Yule–Walker equations to derive the AR(p) model coefficients. The fitted AR(p) model is used to whiten the

* Fax: +1 514 398 3899.
E-mail address: worsley@math.mcgill.ca.
URL: http://www.math.mcgill.ca/keith.
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.02.007

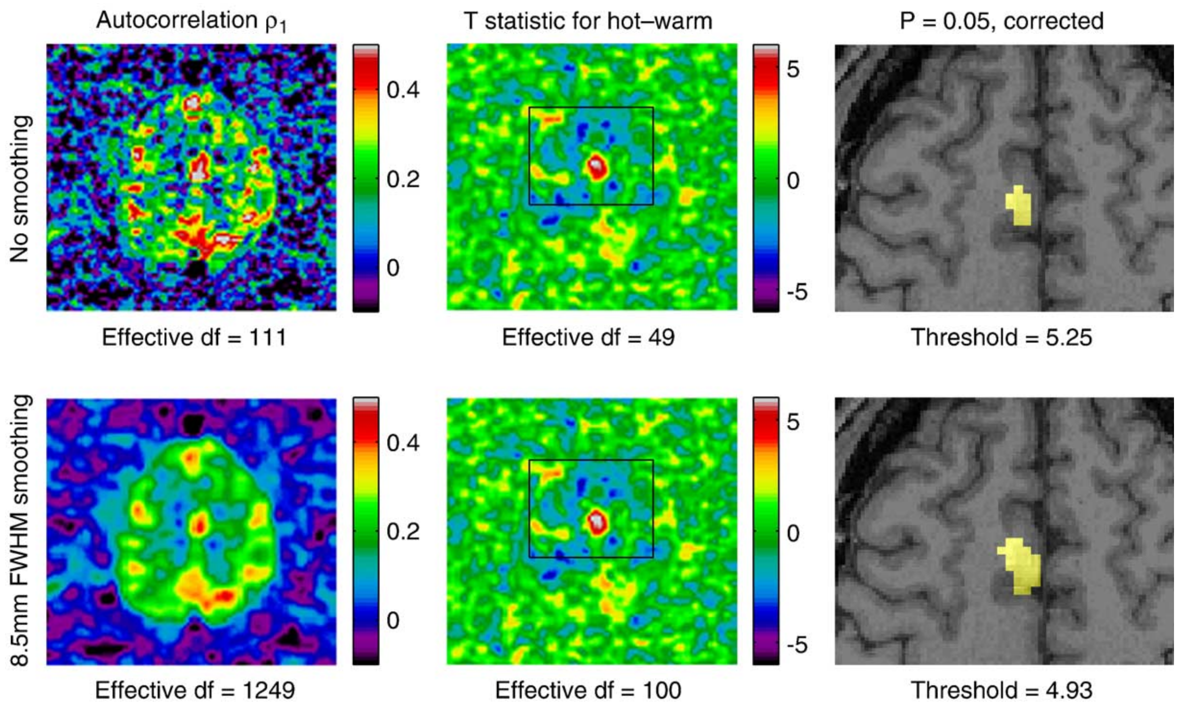

Fig. 1. Temporal autocorrelation (lag 1) of fMRI data from a pain perception experiment, with and without spatial smoothing, the corresponding T statistics for a hot − warm effect, and the detected activation (magnified region framed on the T statistic). Note that with modest smoothing, the effective *df* increases, the resulting $P = 0.05$ threshold decreases, and roughly twice as much activation is detected.

data and the covariates, and the model is re-fitted. Effects of contrasts in the coefficients, and their estimated standard errors S (Eq. (13)), are calculated, leading to T and F statistics, and our statistical inference.

The effect of the spatial smoothing on the effective *df* is as follows. Suppose X is the $n \times m$ design matrix of the linear model whose columns are the covariates, and let c be an m-vector of contrasts in those columns whose effects we are interested in. Let

$$x = (x_1, \ldots, x_n)' = X(X'X)^{-1}c \tag{1}$$

be the least-squares contrast in the *observations*, and let τ_j be its lag j autocorrelation

$$\tau_j = \sum_{i=j+1}^{n} x_i x_{i-j} \Big/ \sum_{i=1}^{n} x_i^2. \tag{2}$$

Let $\mathrm{FWHM}_{\mathrm{data}}$ be the effective FWHM of the fMRI data, and let $\mathrm{FWHM}_{\mathrm{filter}}$ be the FWHM of the Gaussian filter used for spatial smoothing of the temporal autocorrelations. Let

$$f = \left(1 + 2\frac{\mathrm{FWHM}_{\mathrm{filter}}^2}{\mathrm{FWHM}_{\mathrm{data}}^2}\right)^{-D/2} \tag{3}$$

where D is the number of dimensions. Then our main result, proved in Appendix A, is that the effective *df* of the contrast is

$$\tilde{\nu} \approx \nu \Big/ \left(1 + 2f \sum_{j=1}^{p} \tau_j^2\right) \tag{4}$$

where $\nu = n - m$ is the usual least-squares residual *df*. For an F statistic that simultaneously tests k columns of the $m \times k$ contrast matrix C, the effective numerator *df* is the same as Eq. (4) but with x replaced by the normalized matrix of the least-squares contrasts in the observations

$$x = X(X'X)^{-1}C(C'(X'X)^{-1}C)^{-1/2}, \tag{5}$$

so that $x'x$ is the $k \times k$ identity matrix, and with the autocorrelation τ_j replaced by the average of the k temporal autocorrelations of the columns of x.

Temporal correlation of the covariates decreases effective *df*, but since $f \leq 1$, spatial smoothing ameliorates this effect. Reversing this formula (Eq. (4)), we can calculate the amount of smoothing required to achieve a desired *df*. Note that this will depend on the contrast, so we suggest being conservative by taking the maximum of the amounts of smoothing.

Note that we can never get more than the least-squares *df* without smoothing the residual variance as well, in which case the factor f would be applied to all the terms in the denominator of Eq. (4). Of course, we do not wish to this because the residual variance contains too much anatomical structure, and so smoothing could result in serious biases.

The conditions for the result (Eq. (4)) to hold are that the sample size n must be large and the temporal correlations must be small. To see how well this approximation holds up when these conditions are relaxed, we carry out some simulations in the next section.

Results

Simulations

The above theoretical effective *df* (Eq. (4)) is derived under the assumption that the sample size is large and the temporal autocorrelations ρ_j are small. In practice, sample sizes of at least