YNIMG-08805; No. of pages: 11; 4C:

NeuroImage xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Technical Note

The problem of low variance voxels in statistical parametric mapping; a new hat avoids a 'haircut'

Gerard R. Ridgway [a,b,*], Vladimir Litvak [a], Guillaume Flandin [a], Karl J. Friston [a], Will D. Penny [a]

[a] Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, UK
[b] Dementia Research Centre, UCL Institute of Neurology, London, UK

ARTICLE INFO

Article history:
Received 27 August 2011
Revised 7 October 2011
Accepted 11 October 2011
Available online xxxx

Keywords:
SPM
Low variance
VBM
MEG
EEG
Source reconstruction

ABSTRACT

Statistical parametric mapping (SPM) locates significant clusters based on a ratio of signal to noise (a 'contrast' of the parameters divided by its standard error) meaning that very low noise regions, for example outside the brain, can attain artefactually high statistical values. Similarly, the commonly applied preprocessing step of Gaussian spatial smoothing can shift the peak statistical significance away from the peak of the contrast and towards regions of lower variance. These problems have previously been identified in positron emission tomography (PET) (Reimold et al., 2006) and voxel-based morphometry (VBM) (Acosta-Cabronero et al., 2008), but can also appear in functional magnetic resonance imaging (fMRI) studies. Additionally, for source-reconstructed magneto- and electro-encephalography (M/EEG), the problems are particularly severe because sparsity-favouring priors constrain meaningfully large signal *and variance* to a small set of compactly supported regions within the brain. (Acosta-Cabronero et al., 2008) suggested adding noise to background voxels (the 'haircut'), effectively increasing their noise variance, but at the cost of contaminating neighbouring regions with the added noise once smoothed. Following theory and simulations, we propose to modify – directly and solely – the noise variance *estimate*, and investigate this solution on real imaging data from a range of modalities.

Introduction

The statistical parametric mapping (SPM) approach to the analysis of neuroimaging data rests upon the application of frequentist statistics to reject a null hypothesis at a particular voxel, local maximum or contiguous cluster (Chumbley and Friston, 2009; Friston et al., 1994). The null hypothesis, for example of no functional activation or of no group difference in activity or local tissue volume, is commonly tested with a t- or F-contrast of the parameters in a general linear model (Friston et al., 2007). The t-statistic is a signal-to-noise ratio; the significance of the estimated contrast of the parameters is judged with respect to its standard error, which is proportional to the estimated standard deviation of the noise in the model. The F-statistic is a ratio of explained to unexplained variance, which can also be expressed (see Implications for F-contrasts section) as a squared signal-to-noise ratio.

Employing a ratio of signal to noise is necessary because there is no principled parametric method to control the false positive rate when declaring the signal alone to be 'large'. (Worsley et al., 1992) pooled voxels to estimate a spatially stationary noise variance, however, the spatially non-stationary voxel-wise variance estimate proposed by Friston et al. (1991) has been found more appropriate. However, a consequence of each voxel having its own variance estimate is that rejected null hypotheses could relate to unusually low noise variance, as well as or even instead of noteworthy signal.

SPM is intended for smooth (and usually additionally *smoothed*) data, which interacts with this issue, since blurring regions of signal with neighbouring low-variance background regions can cause the significant area to spread into the background, and can shift the peak significance towards the low-variance regions, as observed by Reimold et al. (2006). This reduces the localisation accuracy of the topological features.

Reimold et al. (2006) proposed to address these localisation accuracy problems by returning to consider the underlying signal (the 'contrast' image) within the clusters detected by the conventional t-statistic based procedure. More precisely, significant clusters are grown to accommodate neighbouring voxels with similarly large signal, and the signal itself is visualised in colour-coded maps in place of the usual t-values.[1] However, this modification clearly cannot protect against the unwanted detection of clusters with low signal in regions of even lower variance.

Abbreviations: EEG, electroencephalography; fMRI, functional magnetic resonance imaging; FWHM, full-width at half-maximum; GM, grey matter; MEG, magnetoencephalography; MIP, maximum intensity projection; MNI, Montreal Neurological Institute; ResMS, residual mean squares; SPM, statistical parametric mapping; PET, positron emission tomography; VBM, voxel-based morphometry.

* Corresponding author at: Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, UK. Fax: +44 20 7813 1420.
E-mail address: Gerard.Ridgway@ucl.ac.uk (G.R. Ridgway).

[1] Reimold et al.'s (2006) work resulted in an SPM toolbox for masked contrast images, MASCOI, http://homepages.uni-tuebingen.de/matthias.reimold/mascoi.

1053-8119/$ – see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.neuroimage.2011.10.027

Reimold et al. (2006) considered positron emission tomography (PET) and simulated data. The problem is ameliorated to some extent for functional magnetic resonance imaging. fMRI data typically have lower inherent smoothness and lower applied smoothing (at least for first-level, within-subject, analysis), together with a higher background noise level. However, the problem is far more severe for source-reconstructed magneto- and electro-encephalography (M/EEG). Here, the ill-posed inverse problem requires prior knowledge about the form of the activity in a given task. A commonly used prior is that activation should be spatially sparse, for example with 'multiple cortical sources with compact spatial support' (Friston et al., 2008), which means that low activity – and correspondingly low variance – will be wide-spread even within the grey matter (GM).

For voxel-based morphometry (VBM), the same problem has already been discussed in the literature (Acosta-Cabronero et al., 2008; Bookstein, 2001). (Acosta-Cabronero et al., 2008) proposed that it could be corrected by adding background noise to low-probability voxels in the GM segments. Specifically, following probabilistic tissue classification (Ashburner and Friston, 2005), random noise uniformly distributed between 0 and 0.05 was added to voxels with GM probabilities below 0.05; an approach they termed the 'Haircut' due to its removal of significant voxels outside the skull. Acosta-Cabronero et al. (2008) argued 'intuitively, the statistical effect of noise, with mean and standard deviation an order of magnitude lower than [the probabilities in voxels confidently segmented as GM], being smoothed into GM tissue, can be neglected.' However, they also observed that such a low level of added noise meant that 'the blobs were not completely restored to the glass brain', which leaves open the question of whether a noise-level sufficient to solve the problem fully might have a non-negligible effect on voxels with substantial tissue probability.

The real purpose of adding noise to the data in the Haircut technique is to inflate the error variance σ^2. Changing the data, however, has the unwanted side-effect of altering the estimated parameters and the estimated smoothness, as discussed later. We therefore propose a more incisive modification: that the error variance *estimate* $\hat{\sigma}^2$ (distinguished by the addition of a hat) be directly altered, without requiring any modification of the original data and hence preserving the signal. In brief, we simply add a small value to the estimated error variance. This has only an inconsequential effect in regions with non-trivial signal and variance, but can preclude large statistical values in regions of very low noise, and help to preserve the localisation accuracy of the statistical peaks. This approach is effective and easy to implement; however, it requires us to define what we mean by 'a small value'. In what follows, we evaluate a simple procedure for determining this value automatically. First, we motivate our approach and derive a heuristic using simulated data, then we validate it using real VBM, MEG and fMRI data.

Theory

The main equations related to the contrast c of the parameters β in a linear model of the data in n-vector y (at a particular voxel) with design matrix X (whose Moore–Penrose pseudoinverse is denoted X^+) are:

$$\hat{\beta} = X^+ y \tag{1}$$

$$\hat{\varepsilon} = y - X\hat{\beta} = Ry, \qquad R = I - XX^+ \tag{2}$$

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - rank(X)} = \frac{y'Ry}{tr(R)} \tag{3}$$

$$t = \frac{c'\hat{\beta}}{\hat{\sigma}\sqrt{c'(X'X)^+ c}} \propto \frac{c'\hat{\beta}}{\hat{\sigma}}. \tag{4}$$

Consider the hypothetical scenario of smoothing an infinitesimal 'point source' surrounded by zeros. Both the contrast in the estimated parameters $c'\hat{\beta} = c'X^+y$ and the residual images $\hat{\varepsilon} = Ry$ are linear in the data, so smoothing the data smooths both similarly. The estimated noise standard deviation required for the denominator of the t-statistic is the square root of the residual mean squares (ResMS) $\hat{\sigma}^2 = (y'Ry)/tr(R)$, which is nonlinear in the data, and might appear more complicated. For example, for Gaussian random field data, the estimated standard deviation image would relate to a square root of a Chi-square random field. However, note that because all the residual images have the same spatial profile, this profile is preserved by the squaring and square-rooting operations and by the summation in between them, suggesting that the smoothing will affect the numerator and denominator equally. Supporting this argument, simulations like those described below but with the noise standard deviation tending towards zero, indicate that shape and smoothness of $\hat{\sigma}$ matches that of $\hat{\beta}$ so that the t-map becomes flat and theoretically infinitely extended (see also Chumbley and Friston, 2009). It is therefore clear why low, but non-zero noise, surrounding signals in regions of higher noise, can give rise to the spreading of t-statistic peaks observed in the literature (Acosta-Cabronero et al., 2008; Reimold et al., 2006).

Simulations

To illustrate the nature of the problem and some potential solutions, simple two-dimensional data corresponding to a one-sample *t*-test are simulated. The underlying signal is generated as a point source at the centre (pixel coordinates 20,20) of a 40×40 pixel image, distributed normally with mean 100 and standard deviation 100. A total of $n = 12$ images are simulated, so that the expected t-value of the underlying source is

$$\frac{\beta}{\sigma\sqrt{1/n}} = \sqrt{n} \approx 3.464. \tag{5}$$

The images containing the point source are smoothed with a 10 pixel full-width at half-maximum (FWHM) Gaussian kernel. Similarly smoothed Gaussian noise is added to produce the final data. Two different noise standard deviations are employed: 0.01 and 2; the signal is generated only once, remaining identical for each noise level. Note that because the underlying signal occurs only at one pixel prior to smoothing and that its standard deviation is 50 times higher than that of the high noise level, the high-noise data can also be viewed as an example of applying the Haircut technique of Acosta-Cabronero et al. (2008) to the low-noise data (strictly, the Haircut would not alter the signal pixel itself, but that effect here is trivial).

Fig. 1 (a) and (b) show results for the low- and high-noise data respectively. For the high-noise case, the estimated mean (beta) is very similar to the true value, so its discrepancy is plotted instead (the discrepancy for the low noise case matches that of rows c and d). For the low-noise case, as expected, the estimated noise standard deviation $\hat{\sigma}$ follows the same Gaussian shape as the signal, leading to a t-statistic map with a roughly flat plateau, surrounded by some more erratic values due to boundary effects. For the high-noise case, the t-map plateau is brought closer to the desired shape of the smoothed signal, though at the chosen noise level retains some distortion, with the maximal value displaced from the expected location by about 45% of the applied FWHM. Further increasing the noise level might improve the shape of the t-statistic surface, but at the expense of increasing the errors in the estimated parameter(s) β. There would also be an increasing risk that the non-zero sample mean of the noise itself could actually worsen the shape of the t-map, particularly with low degrees of freedom.

Alternative approaches are investigated in rows (c) and (d) of Fig. 1, each using the original low-noise data, but instead modifying the estimated noise standard deviation image. A reasonable lower

Please cite this article as: Ridgway, G.R., et al., The problem of low variance voxels in statistical parametric mapping; a new hat avoids a 'haircut', NeuroImage (2011), doi:10.1016/j.neuroimage.2011.10.027

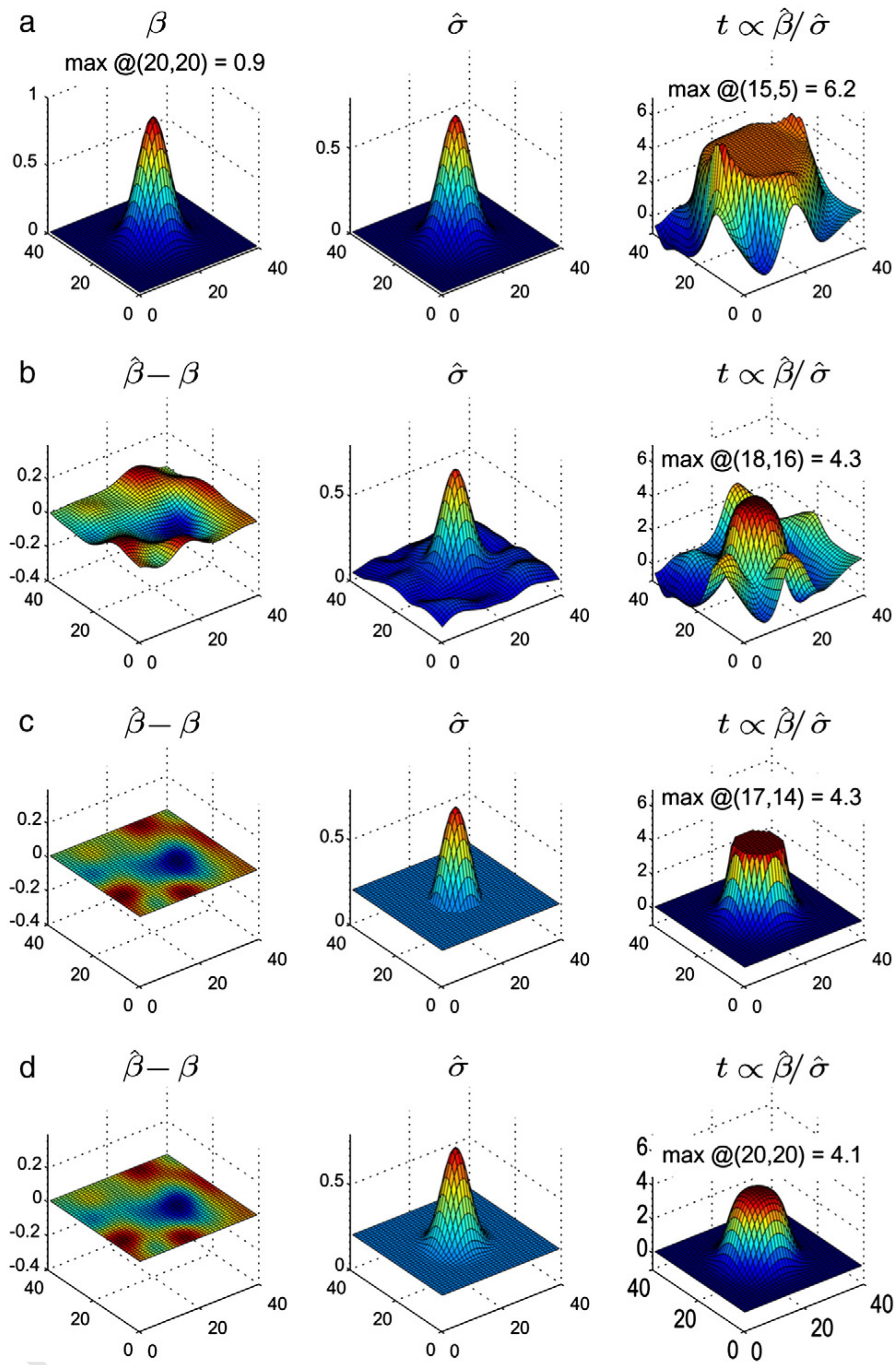

Fig. 1. Simulation results. Columns, from left to right, show: the signal or its error (true signal β in first row; error in estimate $\hat{\beta}-\beta$ in rows b–d); the estimated noise standard deviation $\hat{\sigma}$; the SPM$_t$. Rows correspond to: (a) Low-noise; (b) High-noise (or equivalently, added noise); (c) Low-noise with $\hat{\sigma}$ lower bounded at 0.2; (d) Low-noise with $\hat{\sigma}^2 \to \hat{\sigma}^2 + 0.2^2$.

bound for the (post-smoothing) noise standard deviation is 0.2, and in (c) the estimated standard deviation image is prevented from falling below this bound (values above the bound are left untouched). The original plateau is greatly reduced in diameter, and the Gaussian shape beyond the plateau made more similar to that of the signal. However, the discontinuity introduced to the $\hat{\sigma}$ image results in an undesirable discontinuity in the resultant t-map. Instead of a hard lower bound, therefore, in (d) we propose to modify the estimated noise standard deviation by adding the value of the bound. More precisely, we add the square of the bound to the estimated noise variance, in the expectation that this will impact less upon the pixels which already have suitably high $\hat{\sigma}$ due to the nonlinearity of the square root. That is, we propose a modification of the variance estimate,

$$\hat{\sigma}^2 \to \hat{\sigma}^2 + \delta. \tag{6}$$

Because the data (and hence the parameter estimates $\hat{\beta}$) are not altered, we are able to increase the level of $\hat{\sigma}$ beyond that which could be reasonably achieved with the Haircut technique, obtaining a satisfactory (though still slightly rounded) profile for the t-map, with a correctly located unique maximum value, and without any additional artefacts introduced away from the signal location. Although the addition of δ reduces the maximum t-value compared to using δ as a lower bound, the value of 4.1 is still above that of 3.46 expected for the underlying point source. This is due to the nonlinear effect on $\hat{\sigma}$ from smoothing lower noise regions together with the signal.

It appears that the addition of a small value δ to the estimated noise variance (known as the residual mean squares image in SPM and stored as ResMS.img) is an appealing strategy. However, in this illustration, the lower bound value was chosen by hand to produce satisfactory results. The problem of automatically choosing an appropriate value of δ for typical neuroimaging data is therefore addressed empirically in the following sections.

Material and methods

Three separate modalities are explored: VBM, MEG and within-subject fMRI. In each case, the SPM software (version 8, revision 4290 — but without the modification that is described here) is used to estimate a general linear model and to compute a t-contrast of interest. To illustrate the potential problem at its most severe, the statistical modelling is performed at every non-constant voxel throughout the field of view, i.e. using no explicit mask and no threshold masking. SPM's implicit masking is still used along with the exclusion of voxels that are constant over all scans (which typically excludes only the voxels at the very edges of the field of view that are beyond the six-sigma support of the Gaussian smoothing kernel from any non-zero data). For the MEG data, the source reconstruction process means that the data are zero beyond a moderately tight grey matter mask.

The experimental approach is the same for each modality: the components of the t-statistic – the 'contrast' and the residual mean squares – are displayed; a histogram is used to estimate the distribution of the latter over voxels, and also a joint histogram of the contrast and ResMS, for reasons that will become clear in the results. Note that the joint histogram is only used to determine a suitable procedure from which δ can be estimated from the distribution of ResMS; the eventual (very simple) procedure is not dependent on a particular contrast (and thus can be enacted at the model-estimation stage without requiring any contrasts to be specified). The histograms employ the base-10 logarithm of ResMS; the fact that $\log_{10}(\hat{\sigma}^2) = 2\log_{10}\hat{\sigma}$ obviates the decision of whether to consider ResMS or its square-root. The original t-map is presented alongside the new version using the modified estimate of ResMS ($\hat{\sigma}^2 \rightarrow \hat{\sigma}^2 + \delta$).

VBM data

Structural MRI was obtained from the Open Access Series of Imaging Studies (OASIS) at http://www.oasis-brains.org/. We use the baseline scans from the longitudinal data-set (Marcus et al., 2010), which contains 150 subjects (62 males, 88 females) aged 60to 96. 72 of the subjects were characterized as nondemented throughout the study, 64 were characterized as demented, and 14 subjects were characterized as nondemented at the time of their initial visit but were subsequently characterized as demented at a later visit.

Images were segmented using SPM8's New Segment toolbox (an extension of Ashburner and Friston, 2005) to produce native and 'Dartel-imported' (rigidly aligned to MNI orientation and resampled to 1.5 mm isotropic) segmentations of grey and white matter (WM). Dartel (Ashburner, 2007) was then used to nonlinearly warp all subjects together by simultaneously matching their GM and WM segments to an evolving estimate of their group-wise average (Ashburner and Friston, 2009). The transformations obtained (parameterised by flow-fields) were then applied to the native GM segments together with an affine transformation to MNI space. Probabilistic tissue volumes were preserved ('modulation'). The images were finally smoothed with a Gaussian kernel of 8 mm FWHM.

All 150 subjects were modelled using SPM's flexible factorial design, with a three-level group factor (allowing for unequal variances). Covariates were included to adjust for age, gender and estimated total intracranial volume[2] (Barnes et al., 2010). The contrast of interest tested for reduced GM in the 64 demented subjects compared to the 72 non-demented subjects.

MEG data

We use the multimodal face-evoked responses dataset that is openly available from the SPM website, http://www.fil.ion.ucl.ac.uk/spm/data/mmfaces/, and is described in chapter 37 of the SPM Manual (SPM8 revision 4290). The data are for a single subject, undergoing the experimental paradigm developed by Henson et al. (2003) wherein subjects viewed faces and scrambled images of faces (using random phase permutation in Fourier space). The MEG data were acquired on a 275 channel CTF/VSM system, though one sensor was dropped due to a fault.

The first run (SPM_CTF_MEG_example_faces1_3D.ds) was processed; data were baseline-corrected with baseline between −200 and 0 ms and downsampled to 200 Hz.

Multiple sparse priors (MSP) source reconstruction was performed, which uses a Variational Laplace procedure for automatic relevance determination (ARD), constructing an appropriately sparse solution by selecting from a large number of spatially compact putative sources (Friston et al., 2008).

Standard settings were used (Litvak et al., 2011), with the 'MEG Local Spheres' forward model (Huang et al., 1999), applied to the entire time series. The source power was summarised separately for each trial with a Gaussian window from 150 to 190 ms corresponding to the 'M170' peak in evoked response field (ERF). The power values were smoothed on the cortical mesh using 8 iterations of graph Laplacian smoothing, interpolated from the mesh to a regular three-dimensional volume using a non-linear interpolation method (spm_mesh_to_grid), and then smoothed in 3-D space with a 1 voxel FWHM Gaussian kernel.

A two-sample *t*-test was used to compare 20 trials with faces to 20 with scrambled faces.

fMRI data

The fMRI time-series data is also a standard SPM data-set, available from http://www.fil.ion.ucl.ac.uk/spm/data/face_rep/ and described in chapter 29 of the SPM manual. It consists of a single session of data for one subject from a study of repetition priming for famous and nonfamous faces (Henson et al., 2002). The functional time-series comprises 351 volumes (repetition time 2 s) consisting of 24 descending slices (3 mm thick plus 1.5 mm gap; 64×64 matrix of 3×3 mm^2) of echo planar imaging data (echo time 40 ms). A standard T1-weighted structural MRI is also available.

The data were preprocessed as described in the SPM manual: briefly, the volumes were realigned to correct for head motion, slice-timing discrepancies were corrected, the mean of the realigned functional time-series was coregistered to the structural image, the latter was segmented and the spatial transformation parameters from the unified segmentation (Ashburner and Friston, 2005) were used to spatially normalise the functional images, which were then smoothed with an 8 mm FWHM isotropic Gaussian kernel.

The data were modelled with two conditions, fame (famous face or not) and repetition (first or second presentation), in a 2×2 factorial design. The canonical haemodynamic response with time and dispersion derivatives were used to form regressors from the appropriate stimulus

[2] eTIV, obtained from http://www.oasis-brains.org/pdf/oasis-longitudinal.csv.

Please cite this article as: Ridgway, G.R., et al., The problem of low variance voxels in statistical parametric mapping; a new hat avoids a 'haircut', NeuroImage (2011), doi:10.1016/j.neuroimage.2011.10.027

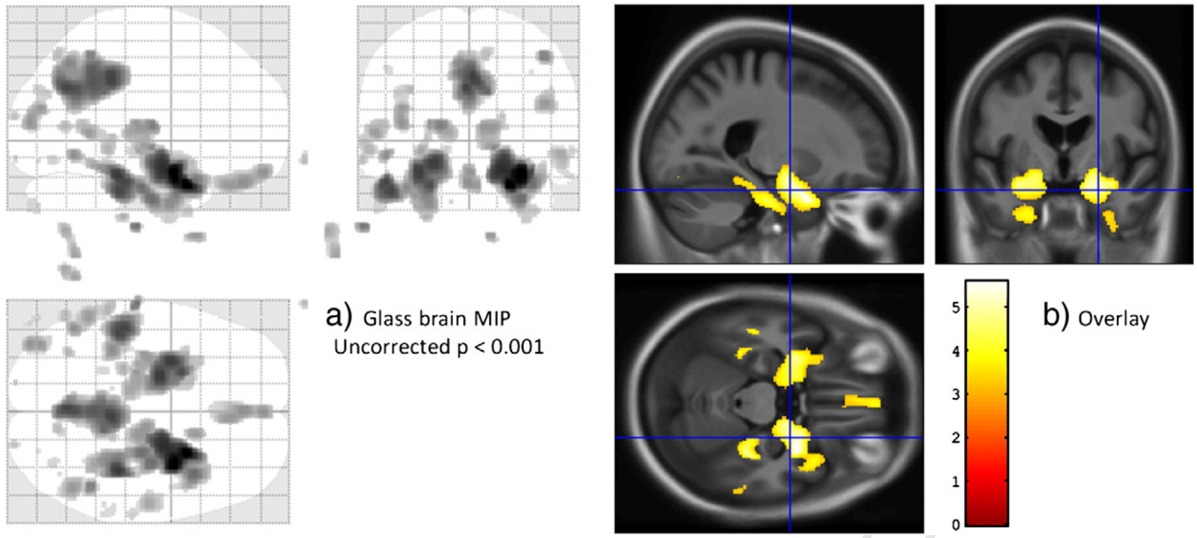

Fig. 2. VBM results. Regions with reduced grey matter in subjects with dementia compared to controls, uncorrected $p<0.001$. (a) 'Glass brain' maximum intensity projection (MIP) showing non-brain false positives (at this uncorrected level) due to low variance outside the brain. (b) Corresponding overlay of thresholded SPM$_t$ on average template, with cross-hairs at the location of maximum 'contrast' value (effect or signal) rather than the more typically reported maximum t-value (signal-to-noise ratio) to provide context for Fig. 3.

functions. Default values were used for other settings (128 s high-pass filter, serial correlations modelled as a first order autoregressive process).

The contrast of interest here is the positive effect of condition on the canonical terms, i.e. the activation in response to faces, averaged over the four cells of the factorial design.

Results

For the purpose of illustration, findings from the VBM data are shown as typically presented in Fig. 2; atrophy in the temporal lobes, posterior cingulate and precuneus is consistent with other reports (Baron et al., 2001; Lehmann et al., 2010), but non-brain regions are visible in the maximum intensity projection (MIP). Next, a series of figures investigating the relationship between the maps of 'contrast' (t-statistic numerator) and of estimated variance are presented, one for each modality studied. Figs. 2(b) and 3(a,c,e,f) show slices at the same location of maximal difference between estimated means for demented and non-demented subjects.

For VBM data, the image of ResMS (Fig. 3 a) is very similar, in terms of the shape of its visibly non-zero regions, to the grey matter of the average template (Fig. 2 b), which reflects the fact that most of the variability (including the unexplained variability that relates to $\hat{\sigma}^2$) in VBM data is in regions with high GM probability. The corresponding panel (a) images for the MEG and fMRI data in Figs. 4 and 5 are strikingly different. For the MEG data, the ResMS image follows the pattern of the contrast image (Fig. 4 c) extremely closely, due to the nature of the MSP source reconstruction method. For the fMRI data, the ResMS is so much higher around the brain stem that its values over the grey matter are visually indistinguishable from the minimum value (white) when the intensity window is set to map the maximum value to black. Changing the intensity window (figure not shown) reveals a broadly uniform pattern over the brain, but with other outlyingly high values, including a very high spot visible near the centre of each slice due to a scanner artefact.

Considering the simple histograms, in Fig. 3(b), clear bimodality is visible, which appears to correspond to distinct foreground and background modes. In contrast, Fig. 4(b) is heavily skewed but shows no pronounced bimodality nor any clear distinction between foreground and background. Fig. 5(b) has a strong unimodal distribution, whose bulk is representative of the values found in grey matter, but with some quite severe outliers at both positive and negative extremes.

The dramatically varying distributions of the voxel-wise variance estimates makes it very difficult to find a suitable generic modelling strategy (attempts to fit mixtures of Gaussians with numbers of components selected by Bayesian model evidence proved unhelpful). This motivates consideration of the joint distribution of variance and signal estimates, in an attempt to find a heuristic background estimate.

The joint histogram for the MEG data (Fig. 4 d) reveals distinct curves; further investigation shows these correspond to separate clusters and are linear when contrast and $\hat{\sigma}$ are considered, thus they appear to represent the simple decay of signal and noise away from the centres of the compact MSP basis functions (see MEG data section). The other joint histograms show similar signal-to-noise relationships evident as the envelope of a much denser pattern, which arises due to the more complicated mixing of many more 'sources' in these modalities.

Based on visual inspection of these and other data-sets' joint histograms,[3] the background estimate or lower bound δ to be added to the ResMS image was chosen to be one thousandth of the maximum value of the ResMS image. Correspondingly, the joint histograms are annotated with a vertical line at the value of

$$\log_{10}\delta = \log_{10}\left(10^{-3} \times \text{maxResMS}\right) = \max(\log_{10}\text{ResMS}) - 3. \qquad (7)$$

This value seems appropriate for both VBM and MEG, but slightly too high for these fMRI data.

The efficacy of the simple modification for VBM and MEG can be seen by comparing the original and modified SPM$_t$ images in Figs. 3 and 4 (e) and (f). For the VBM data, spurious non-brain findings have been dramatically reduced, and there is some evidence of slightly greater anatomical acuity within grey matter. The latter point is more clearly reinforced in the MEG results, where the original SPM$_t$ without modification is unreasonably significant in several regions with very low signal and noise. Reassuringly, the t-values at the

[3] Two further VBM data-sets, two further fMRI data-sets, one further real and one simulated MEG data-set were similarly explored; results were broadly consistent with the exception of the fMRI data-sets, which differed with regard to the severity of their right-tail outliers.

Please cite this article as: Ridgway, G.R., et al., The problem of low variance voxels in statistical parametric mapping; a new hat avoids a 'haircut', NeuroImage (2011), doi:10.1016/j.neuroimage.2011.10.027

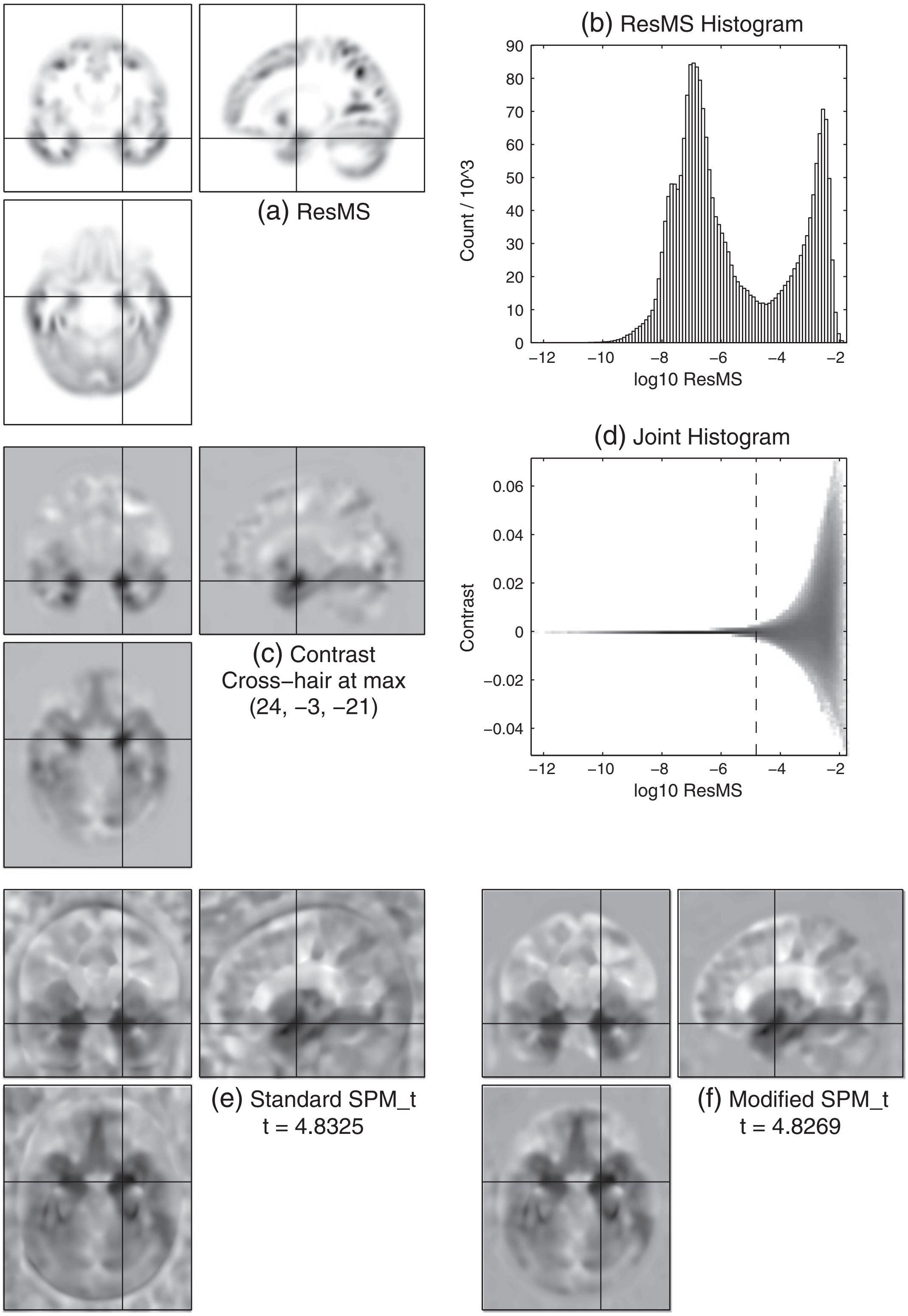

Fig. 3. VBM data. (a) Image of estimated variance or residual mean squares (ResMS). (b) Histogram of log_{10}ResMS over voxels. (c) The contrast of interest, labelled with the MNI coordinates of its maximum value. (d) Joint histogram of contrast value on the vertical axis and log_{10}ResMS on the horizontal axis (which matches that of panel b). The vertical dotted line is located at $log_{10}\delta = max(log_{10}\mathrm{ResMS}) - 3$. (e) Statistical parametric map (SPM$_t$) computed in standard way. (f) SPM$_t$ computed using modified ResMS (with the amount shown dotted in panel d, $\delta = 10^{-3} \times max$ResMS, added on). In panels a and d, white represents zero and higher values or counts are darker; in panels c, e and f, white values are negative and dark values are positive.

location of maximum signal are only reduced by about 0.1%, which should not change their statistical interpretation.

However, for the fMRI data, as could be expected from the location of $log_{10}\delta$ on the joint histogram, the modification has had a more notable effect on the t-value at the location of maximum signal, which has been reduced by over 10%. The t-values changes are more precisely presented in Fig. 6, which plots the changes as a function of the underlying contrast value.

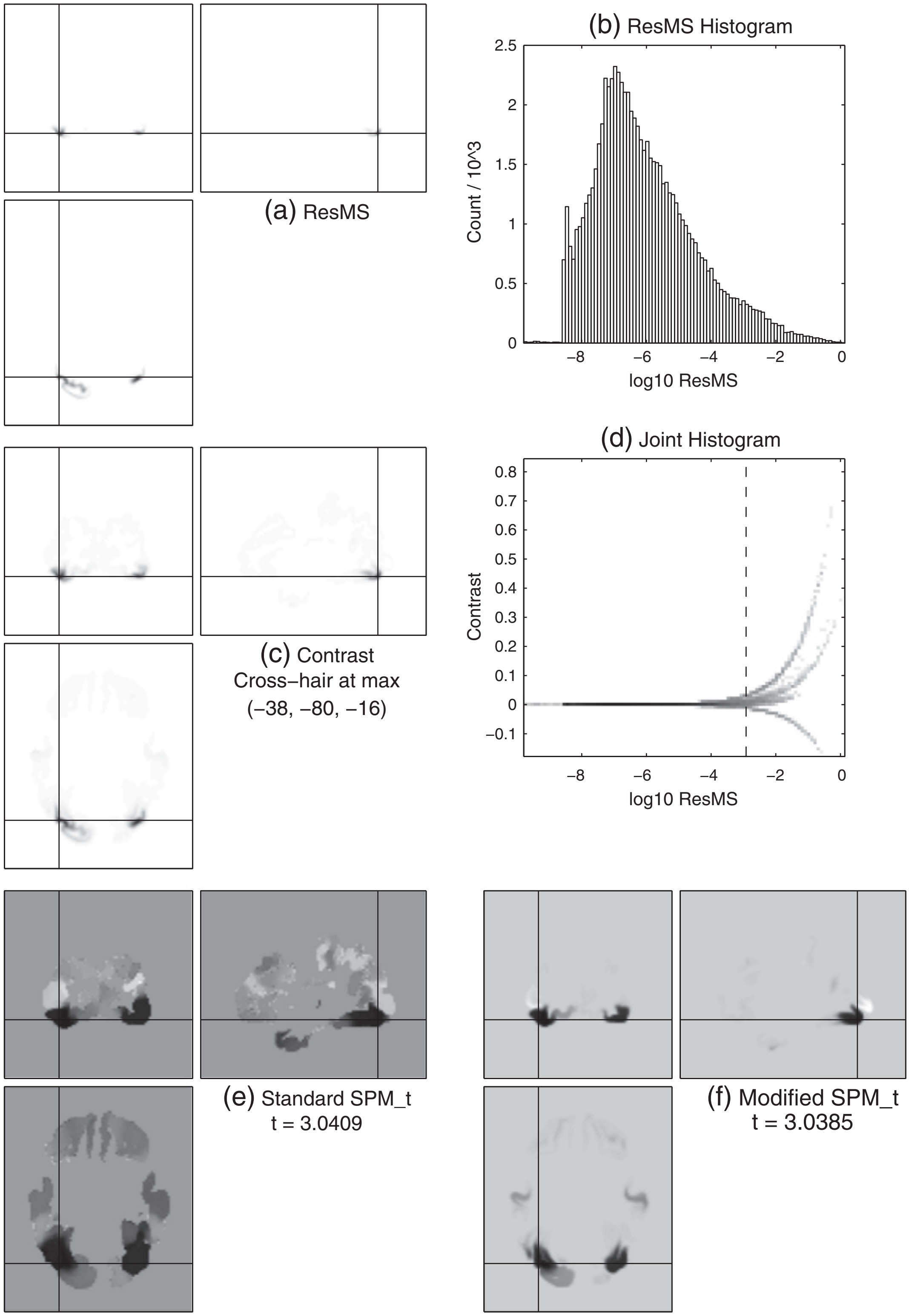

Fig. 4. MEG data, following the same format as Fig. 3.

Discussion

For VBM and MEG data, the problem of low-variance voxels is well addressed by the simple procedure of adding one thousandth of the maximum value to the residual mean squares variance estimate:

$$\delta = 10^{-3} \times \max \hat{\sigma}^2$$
$$\hat{\sigma}^2 \rightarrow \hat{\sigma}^2 + \delta.$$

Source-reconstructed EEG data is expected to behave similarly. However, for the fMRI data here, the reduction in the t-value at the location of maximal signal is over 10%, which is probably unacceptably high, in terms of the consequent reduction in power. This also seems to be true of other within-subject (first-level) fMRI datasets, where artefacts are common but inconsistent in nature, leading to unpredictable behaviour of the joint distribution of signal and noise estimates. Furthermore, the problem of low variance voxels is

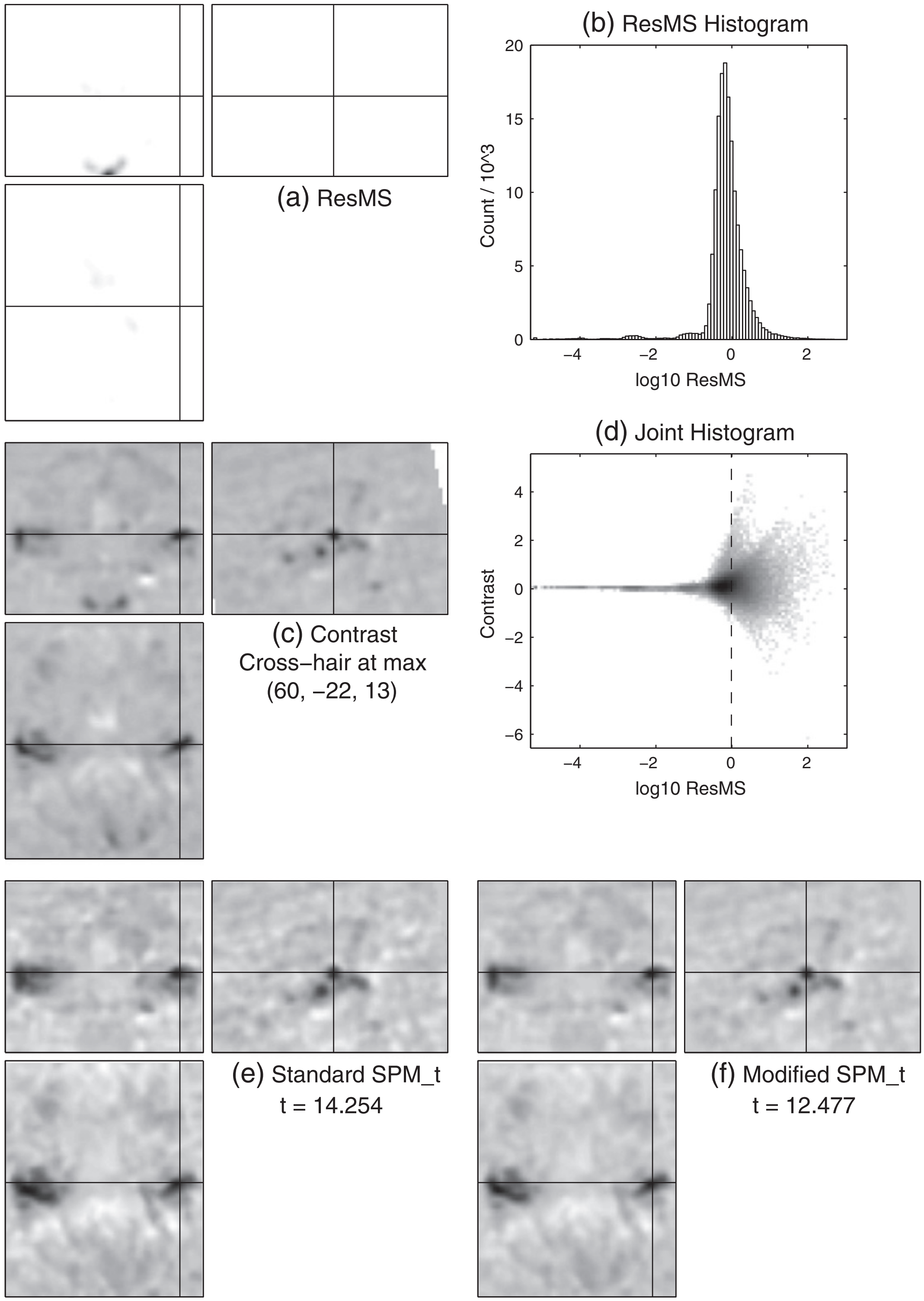

Fig. 5. fMRI data, following the same format as Fig. 3.

itself less severe for within-subject fMRI, due to the smaller amount of smoothing typically applied and the strict masking procedure used by default at the first level.[4]

For between-subject (second-level) fMRI, the analysis mask is the intersection of all the first-level masks, which usually avoids including low-variance non-brain voxels.[5] For these reasons, we do not recommend the procedure here for fMRI.

[4] The masking is not clearly documented, but in `spm_fmri_spm_ui` each image is thresholded at a fraction (by default 80%, defined in `spm_defaults`) of its global value estimated by `spm_global` (the mean of those voxels above one eighth of the original mean), and the intersection of all images' suprathreshold sets defines the overall analysis mask.

[5] If one instead performed the first-level analysis in each subject's native space and normalised the contrast images, then the problem could reappear (e.g. as discussed on the SPM mailing list https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=SPM;ea0f015e.1012 and https://www.jiscmail.ac.uk/cgi-bin/webadmin?A2=SPM;715bcc4e.1012). This is also the reason why SPM's smoothing module has an option to preserve the implicit mask.

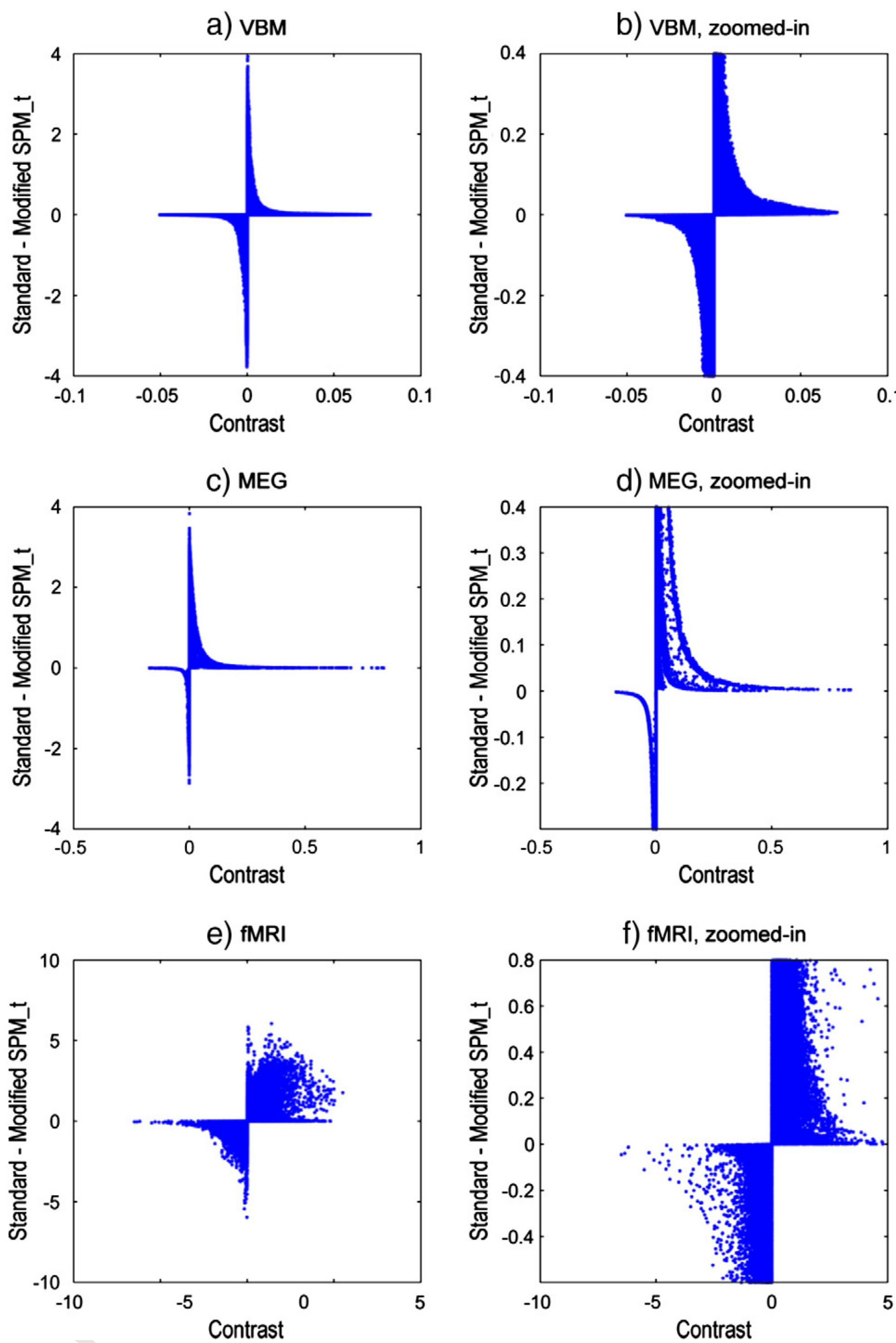

Fig. 6. Comparison of t-statistic values from standard SPM$_t$ and modified SPM$_t$ using the altered ResMS. The difference in t-values (standard–modified) at each voxel is plotted against the contrast value at that voxel. The left column of plots show the full range of values, and the right column zooms in to enable values to be read off more accurately. (a,b) VBM. (c,d) MEG. (e,f) fMRI.

Implications for F-contrasts

Although results have only been presented here for simple t-contrasts, the same low variance problem affects F-contrasts (with single or multiple column contrast matrices C), whose equivalent of Eq. (4) can be written in the form (Christensen, 2002, p. 69):

$$F = \frac{\left(C'\hat{\beta}\right)'(C'(X'X)^{+}C)^{+}\left(C'\hat{\beta}\right)}{\hat{\sigma}^2 rank(C)}, \quad (8)$$

which shows that the same procedure of modifying $\hat{\sigma}^2$ while leaving $\hat{\beta}$ unaltered, also finesses the problem for F-contrasts.

Relation to other procedures

As emphasised by Reimold et al. (2006), the numerator of the t-statistic or contrast image is important because it is probably the most spatially accurate way of locating effects (see also Poldrack et al., 2008 and http://imaging.mrc-cbu.cam.ac.uk/imaging/UnthresholdedEffectMaps). Nevertheless, thresholded SPM$_t$ images remain the most common way of presenting results, and are more closely connected with the assessment of significance. Ultimately, the decision to use the approach of Reimold et al. (2006) or not depends on the particular research question, imaging modality and preferences of the investigator; the new method does not supplant careful consideration of the contrast image.

Due to its implementation prior to smoothing, the Haircut technique proposed by Acosta-Cabronero et al. (2008) can alter the contrast image in regions near to low probability areas (which arguably includes all of the cortex, if the smoothing FWHM is larger than the cortical thickness, as is usually the case), whereas our modification of only ResMS leaves the signal in the contrast image completely intact.

Overly generous masks might also be expected to influence estimation of the smoothness (Kiebel et al., 1999). The resels-per-voxel (RPV) image is typically very low away from the brain (the spatial gradients of low intensity regions tend to be very flat),meaning that the key quantity of interest – the resel count – is only weakly influenced by the inclusion of voxels with very low RPV. However, by adding initially rough noise to non-brain regions, the Haircut technique increases the RPV away from the brain, thus increasing the overall resel count and reducing the power of random field theory based inference. Our modification to ResMS has no impact on RPV estimation.

However, ssq is computed from the residual images themselves (stored temporarily in ResI_* files by spm_spm), not from the (potentially) modified ResMS image, so is unchanged by any modifications to the latter.

Definition of the analysis mask

One might argue that the problem of low variance voxels can be solved simply by defining the analysis mask (Ridgway et al., 2009) to more strictly follow the grey matter, as for example in FSL-VBM (Douaud et al., 2007).[6] This is perhaps partially true for fMRI and VBM, however, as noted by Acosta-Cabronero et al. (2008) and Ridgway et al. (2009), doing so can increase the risk that some true effects will be falsely excluded (particularly in morphometric studies of atrophied or damaged brains). Furthermore, the problem of SPM$_t$ peaks shifting towards regions of lower variance – including those within the brain – would remain, whereas the modification of ResMS proposed here should ameliorate this problem (though admittedly not to the same extent as the method of Reimold et al., 2006). More importantly, with source reconstructed M/EEG data, even a very strict grey matter mask would include some regions of problematically low variance, and attempts to define a very strict signal-based mask might result in too few voxels for the reliable estimation of the smoothness needed for random field theory (Kiebel et al., 1999).

Statistical shrinkage and Bayesian methods

The inflation of the estimated variance proposed here can also be viewed as a shrinkage of the estimated precision towards zero, motivating brief discussion of related statistical shrinkage procedures. The idea of shrinking or deliberately biassing an estimator to improve its performance with respect to some specified loss function was first proposed by Stein (1956) and extended by James and Stein (1961), though Tikhonov had worked on related concepts for integral equations and inverse problems in the 1930s and 40s (Kerimov, 2006). James and Stein (1961) showed that the obvious estimate for the mean of normally distributed samples (assumed to have unit variance for simplicity), which is the maximum likelihood and least-squares estimate, could be improved upon in terms of the estimator's expected squared error by shrinking it towards zero:

$$\bar{y} \to \bar{y} - \frac{n-2}{\bar{y}'\bar{y}}\bar{y} = \left(1-\frac{n-2}{\bar{y}'\bar{y}}\right)\bar{y}. \tag{9}$$

With the goal of estimating covariance matrices, instead of mean vectors, Ledoit and Wolf (2004) derived a shrinkage estimator appropriate for high dimensional problems, which is the asymptotically optimal convex combination of the sample covariance matrix with a scaled identity matrix. This estimate has found application in neuroimaging as part of the 'searchlight' method of Kriegeskorte et al. (2006).

Another related application in imaging is wavelet shrinkage. Bullmore et al. (2004) review methods for denoising and for multiscale spatial hypothesis testing using wavelet shrinkage, in which wavelet coefficients with absolute values below a threshold are zeroed and those above can be either preserved or have the threshold value subtracted from their absolute value, respectively known as 'hard' and 'soft' thresholding.

Bayesian statistics, in which one considers the posterior probability distribution of the aspect(s) of interest can also be used to derive shrinkage procedures. For example, considering β in a linear model to be a random variable having a Gaussian prior distribution with mean μ_β and covariance Σ_β, the maximum a posteriori (MAP) estimate becomes a version of the ML estimate shrunk towards the prior mean, which generalises 'ridge regression' (Friston et al., 2002; Gelman et al., 2003):

$$\hat{\beta}_{MAP} = \left(\sigma^{-2}X'X + \Sigma_\beta^{-1}\right)^{-1}\left(\sigma^{-2}X'y + \Sigma_\beta^{-1}\mu\beta\right) \tag{10}$$

$$= \left(X'X + \sigma^2\Sigma_\beta^{-1}\right)^{-1}\left(X'X\hat{\beta}_{ML} + \sigma^2\Sigma_\beta^{-1}\mu\beta\right). \tag{11}$$

The form of the James–Stein estimator in (9) is clearly closely related to (11) with $\mu_\beta = 0$ and $\sigma = 1$, except that in (9) $\bar{y}'\bar{y} = \hat{\beta}'\hat{\beta}$ depends on the data, while in a conventional Bayesian setting the covariance Σ_β of the prior distribution would not. The notion of empirical Bayesian methods for hierarchical models allows the prior's hyper-parameters to be estimated from the data (Carlin and Louis, 2008; Friston et al., 2002), which can be shown to exactly generalise the James–Stein estimator (Lee, 2004). Similarly, wavelet shrinkage using soft thresholding can also be formulated as a Bayesian procedure with a sparsity-favouring prior over the wavelet coefficients, for example a Laplacian distribution (Bullmore et al., 2004) or a mixture of Gaussians (Flandin and Penny, 2007). Ledoit and Wolf (2004) also present a Bayesian interpretation of their estimator.

It is possible that an appropriate prior distribution for the variability σ^2 could allow something similar to the modification of its estimate proposed here to be derived with an empirical Bayesian approach. This would have the advantage that the amount of modification could be derived from the data itself, instead of being arbitrarily set to some fraction (1/1000 here) of the maximum over voxels, which might conceivably allow the method to adapt more appropriately to fMRI data. However, this would require a somewhat different formulation than the usual hierarchical model in Friston et al. (2002), since the variance estimate becomes a parameter of interest in addition to $\hat{\beta}$, and is therefore left for further work.

Finally, on the topic of Bayesian methods, it is worth noting that the posterior probability mapping (PPM) approach (Friston and Penny, 2003) can entirely circumvent the problem of low-variance voxels undesirably becoming significant: instead of considering significance of each voxel in terms of the probability of the test statistic under the null hypothesis, the Bayesian approach can determine for each voxel the probability that the contrast of its parameters exceeds a specified effect size, and this effect size can be chosen to be non-

[6] http://www.fmrib.ox.ac.uk/fsl/fslvbm/.

Please cite this article as: Ridgway, G.R., et al., The problem of low variance voxels in statistical parametric mapping; a new hat avoids a 'haircut', NeuroImage (2011), doi:10.1016/j.neuroimage.2011.10.027

trivial rather than simply non-zero. For an example of PPM applied to MEG data, see Henson et al. (2007).

Conclusions

For modalities other than fMRI (specifically PET, structural MRI or VBM, and source-reconstructed EEG or MEG), we propose a conservative modification of SPM's residual mean squares image (ResMS) that simply entails adding on 0.1% of its maximum value.[7] It has been shown here that the procedure has very limited effect on regions of meaningfully high signal, while avoiding the problem of artefactually high statistic values in regions with both low signal and low noise.

Acknowledgments

We are grateful for helpful comments from Professor John Ashburner and two anonymous reviewers. This work was supported by the Wellcome Trust, and was undertaken at UCLH/UCL who received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research UK Co-ordinating Centre and has also received equipment funded by Alzheimer's Research UK. The OASIS data were acquired with the support of grant numbers: P50 AG05681, P01 AG03991, R01 AG021910, P50 MH071616, U24 RR021382, R01 MH56584.

References

Acosta-Cabronero, J., Williams, G., Pereira, J., Pengas, G., Nestor, P., 2008. The impact of skull-stripping and radio-frequency bias correction on grey-matter segmentation for voxel-based morphometry. Neuroimage 39, 1654–1665.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113.

Ashburner, J., Friston, K.J., 2005. Unified segmentation. Neuroimage 26, 839–851.

Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. Neuroimage 45, 333–341.

Barnes, J., Ridgway, G.R., Bartlett, J., Henley, S.M.D., Lehmann, M., Hobbs, N., Clarkson, M.J., MacManus, D.G., Ourselin, S., Fox, N.C., 2010. Head size, age and gender adjustment in MRI studies: a necessary nuisance? Neuroimage 53, 1244–1255.

Baron, J.C., Chételat, G., Desgranges, B., Perchey, G., Landeau, B., de la Sayette, V., Eustache, F., 2001. In vivo mapping of gray matter loss with voxel-based morphometry in mild Alzheimer's disease. Neuroimage 14, 298–309.

Bookstein, F.L., 2001. “Voxel-based morphometry” should not be used with imperfectly registered images. Neuroimage 14, 1454–1462.

Bullmore, E., Fadili, J., Maxim, V., Sendur, L., Whitcher, B., Suckling, J., Brammer, M., Breakspear, M., 2004. Wavelets and functional magnetic resonance imaging of the human brain. Neuroimage 23 (Suppl. 1), S234–S249.

Carlin, B.P., Louis, T.A., 2008. Bayesian Methods for Data Analysis, 3 rd edition. Chapman & Hall/ CRC.

Christensen, R., 2002. Plane Answers to Complex Questions: The Theory of Linear Models, 3 rd edition. Springer.

Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited: FDR and topological inference using Gaussian random fields. Neuroimage 44, 62–70.

Douaud, G., Smith, S., Jenkinson, M., Behrens, T., Johansen-Berg, H., Vickers, J., James, S., Voets, N., Watkins, K., Matthews, P.M., James, A., 2007. Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia. Brain 130, 2375–2386.

Flandin, G., Penny, W.D., 2007. Bayesian fMRI data analysis with sparse spatial basis function priors. Neuroimage 34, 1108–1125.

Friston, K.J., Penny, W., 2003. Posterior probability maps and SPMs. Neuroimage 19, 1240–1249.

Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S., 1991. Comparing functional (PET) images: the assessment of significant change. J. Cereb. Blood Flow Metab. 11, 690–699.

Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. Neuroimage 16, 465–483.

Friston, K., Ashburner, J., Kiebel, S., Nichols, T., Penny, W., 2007. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press, Elsevier, London.

Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage 39, 1104–1120.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, 2nd edition. Chapman & Hall/ CRC.

Henson, R.N.A., Shallice, T., Gorno-Tempini, M.L., Dolan, R.J., 2002. Face repetition effects in implicit and explicit memory tests as measured by fMRI. Cereb. Cortex 12, 178–186.

Henson, R.N., Goshen-Gottstein, Y., Ganel, T., Otten, L.J., Quayle, A., Rugg, M.D., 2003. Electrophysiological and haemodynamic correlates of face perception, recognition and priming. Cereb. Cortex 13, 793–805.

Henson, R.N., Mattout, J., Singh, K.D., Barnes, G.R., Hillebrand, A., Friston, K., 2007. Population-level inferences for distributed meg source localization under multiple constraints: application to face-evoked fields. Neuroimage 38, 422–438.

Huang, M.X., Mosher, J.C., Leahy, R.M., 1999. A sensor-weighted overlapping-sphere head model and exhaustive head model comparison for MEG. Phys. Med. Biol. 44, 423–440.

James, W., Stein, C., 1961. Estimation with quadratic loss. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, p. 361.

Kerimov, M., 2006. On the 100th birthday of Academician Andrei Nikolaevich Tikhonov. Comput. Math. Math. Phys. 46, 2027–2030.

Kiebel, S.J., Poline, J.B., Friston, K.J., Holmes, A.P., Worsley, K.J., 1999. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. Neuroimage 10, 756–766.

Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. Proc. Natl. Acad. Sci. U. S. A. 103, 3863–3868.

Ledoit, O., Wolf, M., 2004. A well-conditioned estimator for large-dimensional covariance matrices. J. Multivar. Anal. 88, 365–411.

Lee, P.M., 2004. Bayesian Statistics: An Introduction, 3 rd edition. Hodder Arnold.

Lehmann, M., Rohrer, J.D., Clarkson, M.J., Ridgway, G.R., Scahill, R.I., Modat, M., Warren, J.D., Ourselin, S., Barnes, J., Rossor, M.N., Fox, N.C., 2010. Reduced cortical thickness in the posterior cingulate gyrus is characteristic of both typical and atypical Alzheimer's disease. J. Alzheimers Dis. 20, 587–598.

Litvak, V., Mattout, J., Kiebel, S., Phillips, C., Henson, R., Kilner, J., Barnes, G., Oostenveld, R., Daunizeau, J., Flandin, G., Penny, W., Friston, K., 2011. EEG and MEG data analysis in SPM8. Comput. Intell. Neurosci. 2011, 852961.

Marcus, D.S., Fotenos, A.F., Csernansky, J.G., Morris, J.C., Buckner, R.L., 2010. Open access series of imaging studies: longitudinal MRI data in nondemented and demented older adults. J. Cogn. Neurosci. 22, 2677–2684.

Poldrack, R.A., Fletcher, P.C., Henson, R.N., Worsley, K.J., Brett, M., Nichols, T.E., 2008. Guidelines for reporting an fMRI study. Neuroimage 40, 409–414.

Reimold, M., Slifstein, M., Heinz, A., Mueller-Schauenburg, W., Bares, R., 2006. Effect of spatial smoothing on t-maps: arguments for going back from t-maps to masked contrast images. J. Cereb. Blood Flow Metab. 26, 751–759.

Ridgway, G.R., Omar, R., Ourselin, S., Hill, D.L.G., Warren, J.D., Fox, N.C., 2009. Issues with threshold masking in voxel-based morphometry of atrophied brains. Neuroimage 44, 99–111.

Stein, C.J., 1956. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Probab, pp. 197–206.

Worsley, K.J., Evans, A.C., Marrett, S., Neelin, P., 1992. A three-dimensional statistical analysis for CBF activation studies in human brain. J. Cereb. Blood Flow Metab. 12, 900–918.

[7] Software note: The modification has been released as part of revision 4290 (04-Apr-2011) of SPM8, available from http://www.fil.ion.ucl.ac.uk/spm/software/spm8/#Updates.

Please cite this article as: Ridgway, G.R., et al., The problem of low variance voxels in statistical parametric mapping; a new hat avoids a ‘haircut’, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.10.027