Visualizing Variance with Percent Change Threshold

Thomas Nichols

Department of Biostatistics, University of Michigan

Ann Arbor, MI 48109, U.S.A.

nichols@umich.edu

Phone: 734-936-1002, Fax 734-763-2215

December 19, 2002

Abstract

This report proposes a standard method to characterize variability and sensitivity of functional neuroimaging experiments. Most functional neuroimaging analyses result in statistic images of signal-to-noise. However, it is essential to examine both signal and noise individually, to characterize sensitivity and to identify artifacts. Confidence intervals simultaneous convey signal and noise but are poorly suited to imaging. We describe an image which conveys the same information as a confidence interval when used with an image of estimated signal. This image is the half-length of a two-sided $(1-\alpha)100\%$ confidence interval, and is most easily seen as the minimum change required to obtain a significant activation at level α (either corrected or uncorrected). If the estimated signal has units of percent change, then we call this image a level α percent change threshold (PCT). We demonstrate the method on two fMRI datasets, showing that PCT can vary by an order of magnitude across the brain.

Keywords: Confidence Intervals, global estimation, mode estimation, statistical inference, statistical power.

Running title: Visualizing Variance

1 Introduction

The standard approach to functional neuroimaging data analysis consists of fitting a model at each point in the brain. Evidence for an effect of interest is typically summarized in a single t or p value. While these are useful summaries of reliability or signal-to-noise, it is important to examine signal and noise individually. Inspection of images of estimates and their standard deviation is useful, but if not appropriately scaled, they can be difficult to interpret.

In general statistical applications, confidence intervals are used to show an estimate (the signal) and its variability (the noise) simultaneously. A confidence interval also conveys the results of a hypothesis test: If the null-hypothesized value is outside the interval, the test rejects the null. In imaging, where we struggle to present three or more dimensions on two-dimensional displays, there is no obvious way to present confidence intervals at every voxel.

In this technical report we describe the transformation of images of standard deviation into images of the half-length of a two-sided $(1-\alpha)100\%$ confidence interval; this measure is most easily described as the minimum change required to obtain significant activation at level α (either corrected or uncorrected). If the change images have units of percent change, then we call this image a level α percent change threshold (PCT). A PCT image is a scalar multiple of either a standard deviation image or a coefficient of variation image and hence is easily created.

While the PCT image implements a confidence interval when viewed with an image of signal, it is also useful on its own. Aside from characterizing possible artifacts, regions of increased PCT indicate lower power and possible Type II errors. Further, this reduced power is precisely described in units of percent change—how large a signal one would require to obtain significance.

The method is very general and applies to any model that produces a t ratio. The only restriction is that a unit change in the numerator of a t statistic corresponds to a unit change in the data. We give explicit guidelines on how to enforce this for linear models. Matlab software is available to create PCT images (`http://www.sph.umich.edu/~nichols/PCT`).

While others have recommended the use of coefficient of variation images (Glover & Lai, 1998) for fMRI, we believe this is the first use of scaling such images to yield the concrete interpretation just described.

In the next section we first introduce the approach for an arbitrary model, then examine several

considerations in its use. Next we describe its form for the general linear model, with independent and dependent data, and then consider two special cases. We demonstrate the method on two fMRI datasets. In an appendix we describe a novel approach to estimating global image intensity.

2 Methods

Consider models where the statistical evidence a single voxel can be summarized by a t statistic:

$$t = \frac{\hat{\delta}}{\text{sd}(\hat{\delta})} = \frac{\hat{\delta}}{\text{nsd}(\hat{\delta})\hat{\sigma}}, \tag{1}$$

where $\hat{\delta}$ is the summary measure of the effect of interest, $\text{sd}(\hat{\delta})$ is the standard deviation of that estimate, and $\text{nsd}(\hat{\delta}) = \text{sd}(\hat{\delta})/\hat{\sigma}$ is the normalized standard deviation. We make no assumptions here, other than that $\hat{\sigma}$ can be factored out as per $\text{nsd}(\hat{\delta})$ (though a modification easily accommodate this).

A confidence interval is obtained by inverting the corresponding hypothesis test (Schervish, 1995):

Hypotheses	Rejection Region	Confidence Interval
$\mathcal{H}_0 : \delta = 0$ $\mathcal{H}_A : \delta \neq 0$	$(-\infty, -T_\alpha] \bigcup [T_\alpha, \infty)$	$\left[\hat{\delta} - T_\alpha \text{nsd}(\hat{\delta})\hat{\sigma},\ \hat{\delta} + T_\alpha \text{nsd}(\hat{\delta})\hat{\sigma}\right]$

where T_α is an activation threshold for a two-sided level α test. T_α may simply be the $\alpha/2$ critical value of a t distribution, or it may be larger to account for the multiple comparisons problem.

If 0 lies outside of the confidence interval we reject $\mathcal{H}_0$and declare $\hat{\delta}$ significant; equivalently if $\hat{\delta}$ exceeds $T_\alpha \text{nsd}(\hat{\delta})\hat{\sigma}$ in magnitude we reject $\mathcal{H}_0$ (the same result holds for one-sided hypotheses, though T_α will be different). Hence, $T_\alpha \text{nsd}(\hat{\delta})\hat{\sigma}$ is a statistical threshold *in units of* $\hat{\delta}$. If the images were quantitative we could stop here and use the following result: Scale standard deviation by $T_\alpha \text{nsd}(\hat{\delta})$ to obtain images of minimum detectable change.

In functional neuroimaging, however, images are typically nonquantitative, and so we instead standardize by grand mean or baseline μ. A confidence interval for percent change is

$$\left[\frac{\hat{\delta}}{\mu} - T_\alpha \text{nsd}(\hat{\delta})\frac{\hat{\sigma}}{\mu},\ \frac{\hat{\delta}}{\mu} + T_\alpha \text{nsd}(\hat{\delta})\frac{\hat{\sigma}}{\mu}\right] 100\%.$$

Thus the statistical threshold for percent change is:

$$\text{PCT} = T_\alpha \text{nsd}(\hat{\delta}) \, \frac{\hat{\sigma}}{\mu} \, 100\%. \tag{2}$$

This result can be read: Scale coefficient of variation by $T_\alpha \text{nsd}(\hat{\delta})$ to obtain images of minimum detectable change. For either result, the key observation is thus: When appropriately scaled, either standard deviation or coefficient of variation can be interpreted as a detection threshold.

2.1 Considerations

While this result appears to be just a simple manipulation of a confidence interval, in fact there are several important subtleties. We need to choose an activation threshold T_α and determine the appropriate baseline when applying this result voxel-by-voxel. The foremost concern, however, is the PCT's lack of scale invariance.

2.1.1 Invariance & Units

A t ratio is unitless and invariant to scaling of the statistic. In contrast, a confidence interval has the units of the estimate $\hat{\delta}$ and hence the PCT is not scale invariant. For example, if $\hat{\delta}$ is redefined to equal double the response magnitude, the t statistic will not change, but the PCT would double (via the $\text{nsd}(\hat{\delta})$ term). So care must be taken to ensure that the estimates have the correct units. In particular, the statistic must be constructed such that a unit change in $\hat{\delta}$ corresponds to a unit change in the data. We return to this below when we consider specific models.

2.1.2 Activation Threshold

Another consideration is the choice of activation threshold T_α, in particular whether an uncorrected or corrected threshold should be used. While use of a uncorrected threshold yields the interpretation "How large of a change is needed to attain significance", the corrected threshold adds the qualifier "accounting for the search over all brain voxels". We see two different uses for each.

If there is no specific *a priori* hypothesis, then a corrected threshold should be used. This offers a type of sensitivity analysis, conveying the minimum change detectable when searching the whole brain for activations. For diagnosis purposes, however, an uncorrected threshold may be

more appropriate. If an artifact is identified, we might ask "Even *if* we had identified this problem voxel *a priori*, what is the minimum detectable percent change here?" An uncorrected threshold gives an assessment of the intrinsic sensitivity independent of the number of voxels examined. In practice, since the difference between the two choices is just a scale factor, we actually prefer to use both, labeling a color scale with the values from both an uncorrected and corrected threshold.

To account for multiple comparisons, any method can be used which produces a corrected t threshold. We use the False Discovery Rate (FDR) to account for the multiple comparisons problem (Genovese *et al.*, 2001). A $q = 0.05$ FDR threshold controls the expected proportion of false positives *among* suprathreshold voxels. This is in contrast to a standard familywise error method, which controls the chance of obtaining *any* false positives.

2.1.3 Baseline

When equation (2) is applied voxel-by-voxel we may want to consider using a common, global value for μ. The motivation for using a global μ is that it allows us to simply scale a standard deviation image into interpretable units. For fMRI T2* images, gray matter and white matter voxels are roughly homogeneous, and hence a pooled μ seems reasonable and is easily obtained. For PET blood flow images, there is considerable gray-white contrast, and hence a pooling μ only over gray matter may be more appropriate.

For estimating a global μ we find the mode of the baseline image more satisfactory than the mean or, say, the median. The mean is affected by outliers in the image, such as hyperintense CSF voxels, and both the mean and median depend strongly on a brain-nonbrain threshold. In particular, we find that a simple global mean estimate (e.g. that used in SPM99, `http://www.fil.ion.ucl.ac.uk/spm`) consistently under estimates typical gray matter intensity (see Figure 5). In Appendix A we provide a simple, robust method for estimating the mode.

2.1.4 Visualization

While the PCT image is useful alone, it is most useful when paired with a percent change image. To take best advantage of this we recommend that percent change and PCT be displayed side by side with an orthogonal slice viewer which prints the intensities of the selected voxel.

With such a set up, $(1-\alpha)100\%$ confidence intervals can read off as the percent change plus or minus the PCT value. This is the tool we use in the results section, and is also available at `http://www.sph.umich.edu/~nichols/PCT`.

2.2 Specific Models

We now consider the details for application to the General Linear Model and two special cases, a two sample t-test and a correlation.

2.2.1 General Linear Model, Independent Data

The GLM at a single voxel is written

$$Y = X\beta + \epsilon \tag{3}$$

where Y is a n-vector of data, X is a $n \times p$ matrix of p predictors, β is a p-vector of unknown, fixed parameters, and ϵ is a n-vector of unknown, random errors. First assume that the errors are uncorrelated with homogeneous variance ($\mathsf{Var}(\epsilon) = I\sigma^2$), but we later show how the results generalize to other cases.

The most important requirement for this work is that the design matrix be scaled appropriately. That is, the columns must be constructed such that a unit change in a β results in a unit change in Y. In PET and fMRI with box-car covariates, this is a simply a matter of using 1's and 0's to construct X. For continuous covariates, this amounts to ensuring that the baseline-to-peak difference is unity. Note this is not always equivalent to scaling the max-to-min range to be unity: In fMRI, the minimum value may occur in a post-stimulus overshoot, and not at baseline.

Our summary measure is some linear combination of the parameters, specified with a contrast c, a row vector of length p:

$$\hat{\delta} = c\hat{\beta} = c(X'X)^{-1}X'Y \tag{4}$$

The normalized standard deviation of this estimate is

$$\text{nsd}(\hat{\delta}) = \sqrt{c(X'X)^{-1}c'}. \tag{5}$$

Hence the PCT is

$$T_\alpha \sqrt{c(X'X)^{-1}c'}\ \frac{\hat{\sigma}}{\mu}\ 100\%. \tag{6}$$

Just as the covariates must be constructed with care, the contrasts must be constructed to preserve the units of the predictors in the design matrix. The sufficient condition is that the sum of the positive elements is one and the sum of the negative elements is -1. For all non-negative or all non-positive contrasts, this can be simply seen as ensuring that the contrast expresses an average. Enforcing this constraint individually on positive and negative elements amounts to checking that the contrast is a difference of averages. This requirement is easily enforced in software and can be done, say, when estimability checked.

If a model has already been estimated with an inappropriately scaled X matrix, it is easy to correct the scaling with the contrast. Let D be a diagonal matrix such that XD is appropriately scaled, then $D^{-1}\beta$ will yield parameters with the correct units. Use of the contrast $c_{\text{adj}} = cD^{-1}$, where c has an absolute sum of unity, will scale the betas as desired without refitting the model.

The preceding constraints ensure that $\hat{\delta} = c\hat{\beta}$ has the same units as the data. Note then, that $(c\hat{\beta}/\mu)100\%$ is the percent change. Moreover, if the grand mean μ is homogeneous across the brain and equal to 100, the contrast images $c\hat{\beta}$ can be interpreted as approximate percent change images.

If "second level", random effects models are used (Holmes & Friston, 1999), one only needs to ensure that each individual's contrast is appropriately constructed (and that all subjects had a common global baseline intensity). Then, as long as the second level contrast is constrained as specified above, the resulting contrast image will have the appropriate units. Percent change images could be constructed by creating an intersubject mean baseline image, or, if subject's baseline images were homogeneous and 100, the second level contrast image could serve as an approximate percent change image.

One final consideration is the selection of the contrast when there is more than one possible contrast. If there are K contrasts, one possibility would be to take the root mean square of the normalized variances,

$$\sqrt{\frac{1}{K}\sum_k c_k(X'X)^{-1}c_k'}. \tag{7}$$

Another would be to use the contrast with the largest normalized standard deviation. We favor this second, more conservative option.

2.2.2 General Linear Model, Dependent Data

The PCT generalizes to dependent data, and it is straightforward to show that when either prewhitening or precoloring is used the PCT is

$$T_\alpha \sqrt{c(KX)^- KVK'((KX)^-)'c'}\ \frac{\hat{\sigma}}{\mu}\ 100\% \tag{8}$$

where K is the (de)correlating matrix, $V = \mathsf{Var}(\epsilon)/\sigma^2$ is the intrinsic autocorrelation, and superscript minus denotes the pseudo inverse (Graybill, 1976, pg. 28). In prewhitening K is chosen such that KVK' is an identity matrix, while in precoloring K chosen such that $KVK' \approx KK'$.

Note that because the matrix K is applied to both the data and the design matrix, there are no additional invariance issues—the betas' units are unaltered by K.

2.2.3 Two Sample t Test

It is useful to consider special cases to build intuition. Consider a two sample t-test, where n_1 scans are compared to n_2 scans; the design matrix is $n \times 2$ ($n = n_1 + n_2$) and consists of one column with n_1 1's indicating the scans in the first condition, the second column with n_2 1's indicating the second condition.

Note that our design matrix is appropriately scaled: A unit change in either β_1 or β_2 will result in a unit change in Y. To preserve these units we use the contrast $c = [-1/2\ 1/2]$. Plugging these into equation (6) yields the PCT

$$T_\alpha \frac{1}{2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\ \frac{\hat{\sigma}}{\mu}\ 100\%. \tag{9}$$

If the two groups are balanced, the factor reduces to

$$\frac{T_\alpha}{\sqrt{n}}\ \frac{\hat{\sigma}}{\mu}\ 100\%. \tag{10}$$

Observe that, for this simple case, if our data have a spatially homogeneous baseline image with mean 100, then scaling standard deviation images by $T_\alpha/\sqrt{n}$ yields PCT images.

2.2.4 Correlation

Next consider a simple correlation. Again a $n \times 2$ design matrix is used to represent this case. Let the first column of X be a column of ones; let the second be the covariate, say x. As described

above, the predictor x must be constructed such it represents a unit response; if we were using a sine wave, we would want to scale it by 1/2 so that maximum to minimum difference is unity; if it were a canonical HRF, we would ensure that the baseline to activation plateau range was one. The appropriate contrast is then $c = [0\ 1]$, and this yields the factor

$$\frac{T_\alpha}{\sqrt{n}} \frac{1}{\sqrt{\frac{1}{n}\sum_j (x_j - \bar{x})^2}} \frac{\hat{\sigma}}{\mu} 100\%. \tag{11}$$

This is the same as equation (10) except for the term involving $\frac{1}{n}\sum_i (x_i - \bar{x})^2$, the second moment of the covariate. This reveals that by making our design as variable as possible, with many ON's and OFF's, we can detect smaller changes—again corresponding to intuition. This interpretation, of course, is subject to the constraint that the covariate has a unit response scaling and, of particular concern for fMRI, that linearity holds and that the data are independent.

In some cases the covariate will have units of its own that will have to be taken into account. For example, a study involving subjective assessment may use a visual analog scale (VAS) that is scored 0 to 10. If VAS score is entered into the design matrix as a covariate, a percent change image will have the interpretation of percent signal change per unit VAS increase, and a PCT image must be similarly interpreted.

Irregularly-spaced, event-related fMRI presents some unique aspects as well. There may be an isolated event represented by the HRF with unit magnitude, and there may also be several closely spaced events that combine to yield a magnitude much larger than unity. In this case, there is no scaling necessary, and a percent change image based on this design will have the interpretation of percent change in response to an isolated event.

2.3 Data

We demonstrate this method with two datasets. For each we create two PCT images, one using the voxel-wise baseline, and one using a global baseline; a global baseline was estimated with the mode of the mean image after discarding voxels less than the antimode (See Appendix A). General agreement between these two images will validate the use of a global value. By using a color scale with dual axis labels, we simultaneously view both $\alpha = 0.05$ uncorrected and the $q = 0.05$ FDR-corrected PCT images.

The first study was a visual stimulation task, with conditions of fixation, right hemifield stimulation, and left hemifield stimulation. Fixation lasted 3 seconds and each stimulation condition lasted 5 seconds; they alternated as Fix-Left-Fix-Right-Fix-Left, etc. The PCT was based on the Left versus Right contrast. A single acquisition was collected consisting of 160 images, with a TR of 2000ms, a TE of 30ms and a FA of 80°. Images were 64×64 with 30 slices, with voxel dimensions of 3.125×3.125×4.0mm (no skip). A single-shot spiral sequence was used.

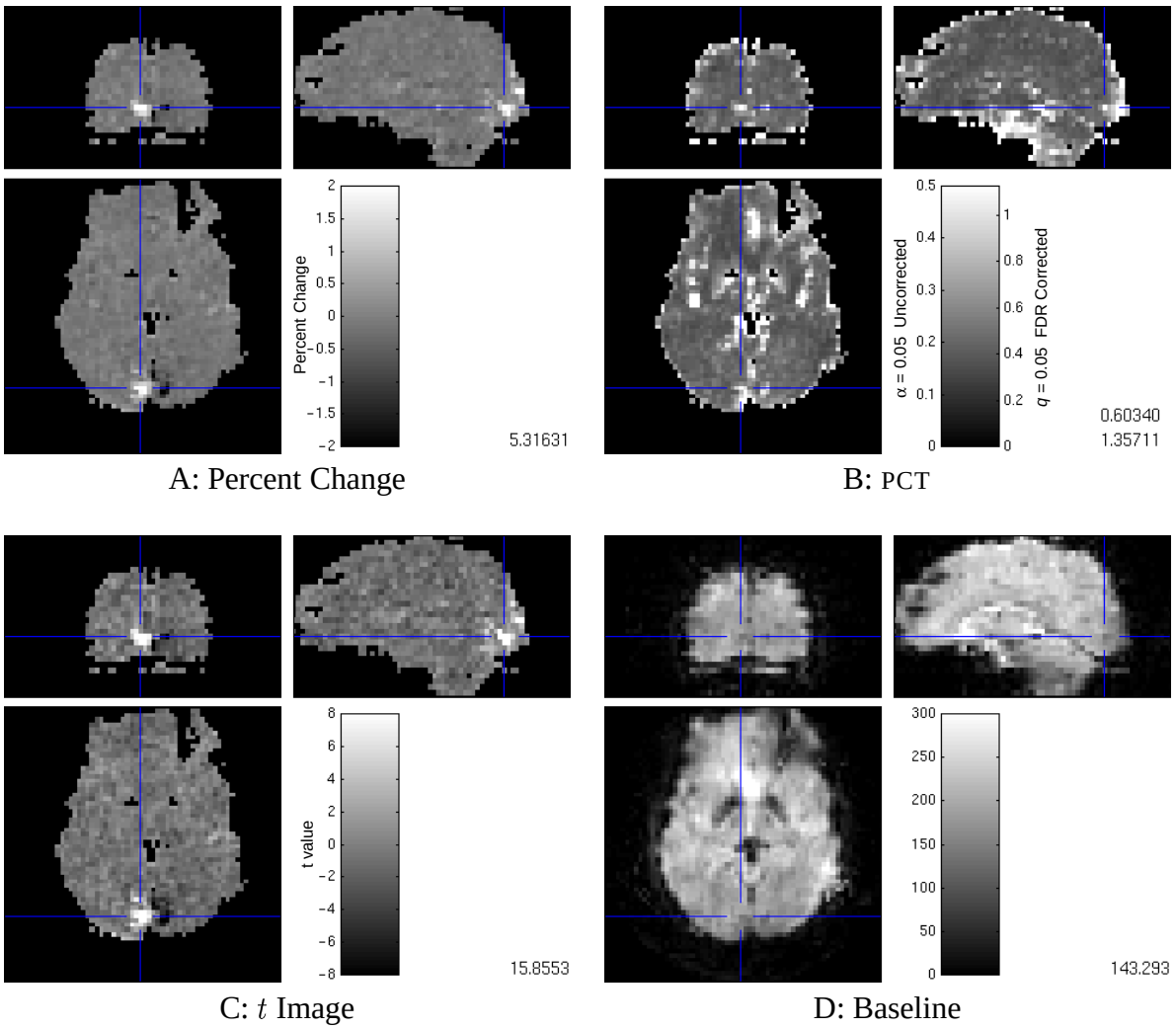
The second study consisted of a visual presentation task. The study employed a block design paradigm with three types of conditions. Presentation with (1) emotionally aversive, (2) emotionally neutral or (3) blank pictures. The design was balanced, so all pairwise contrasts had the same standardized variance. There were four acquisitions, each having three blocks of each condition in pseudo-randomized order; each acquisition had 90 scans collected with a TR of 3000ms, a TE of 15ms, and a FA of 65°. Images were 64×64 with 28 slices, with voxel dimensions of 3.125×3.125×4.0mm (no skip). A two-shot spiral sequence (Noll *et al.*, 2000) was used to avoid susceptibility-related image distortions.

Both datasets were collected on a 3T GE Signa magnet, and each were corrected for slice timing effects and subject motion using SPM99 (`http://www.fil.ion.ucl.ac.uk`). The analysis mask was determined by thresholding the mean image at the antimode (See Appendix A).

3 Results

Figure 1 shows the results of the hemifield stimulation study in the recommended format of percent change paired with PCT; for comparison and reference, the t and baseline images are also shown. While the t and percent change images look similar, the percent change image has readily interpretable units. Referencing the percent change image to the PCT image indicates if a given voxel is significant. The voxel of interest exhibits a 5.316% change, which is larger than the PCT 1.357% ($q = 0.05$ FDR corrected), and hence the voxel is significant. Equivalently, we immediately have a 95% confidence interval, (3.959%, 6.673%). The percent change and t images for the second study are qualitatively similar to these and hence omitted.

Figures 2 and 3 show the utility of the PCT image for characterizing expected and unexpected artifacts. Figure 2 shows the PCT image for the second study. The t statistic image (not shown)



A: Percent Change B: PCT

C: t Image D: Baseline

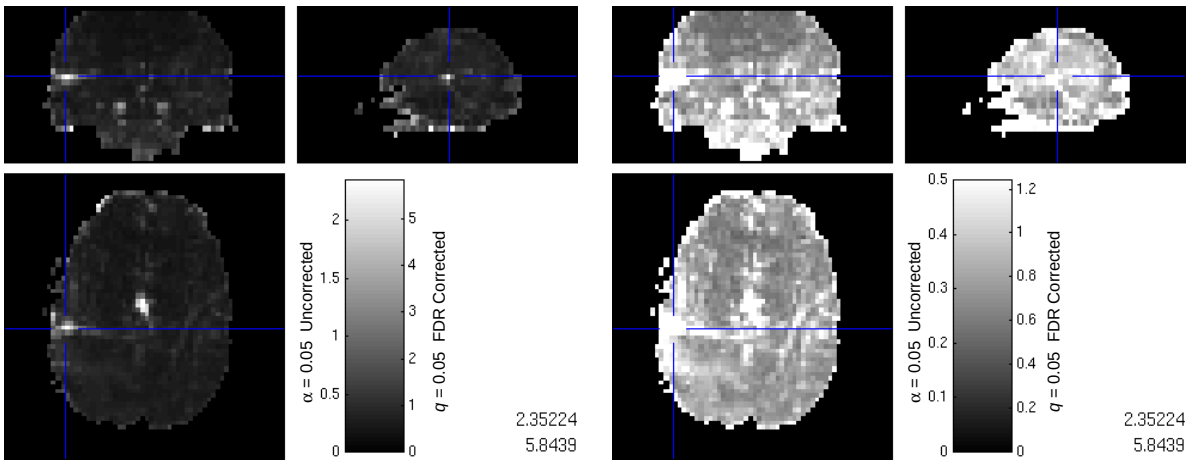
Figure 1: Results of the first study. Images are displayed with orthogonal sections through the voxel of interest; numbers in the lower right of each sub-figure are the value of the voxel of interest. The percent change (A) and t image (C) appear similar, but the former is more interpretable: The voxel under the crosshair has a t of 15.1, which simply reads as "very significant"; in contrast, the percent change image shows that the voxel exhibits a 5.316% change. The PCT image (B) shows that that a 95% confidence interval at that voxel is $5.316 \pm 1.357\%$ (FDR-corrected $q = 0.05$; uncorrected $\pm 0.603\%$; note the dual axis colorbars). The confidence interval doesn't cover zero and hence this voxel is significant. Further, the homogeneity of the PCT image reveals that, typically, we have a precision of about $\pm 0.6\%$ (corrected). The baseline image (D) is provided for reference.

revealed nothing unusual, but the PCT and standard deviation images had evidence of artifacts (likely due to physiological effects or head motion). Figure 2-A and 2-B shows the PCT image for two different intensity windowings; Figure 2-B was windowed such that the modal value was approximately in the middle of the color bar. Hence we can immediately see that, in most of the brain, we can detect changes as small as 0.6% when accounting for multiple comparisons (0.25% uncorrected); however, in the left temporal-parietal region around the artifact, the PCT ranges from 2.5% to over 5% (corrected). Hence there could be a signal on left that exceeds one on the right by a factor of four, and yet only the signal on the right will be detected. This would be of profound importance if laterality effects were of interest in this study.

(As a practical matter, the best course of action for this data would be to identify the images causing the artifact and remove them from the analysis. We are developing methods to efficiently localize such problems (Nichols & Luo, 2001)).

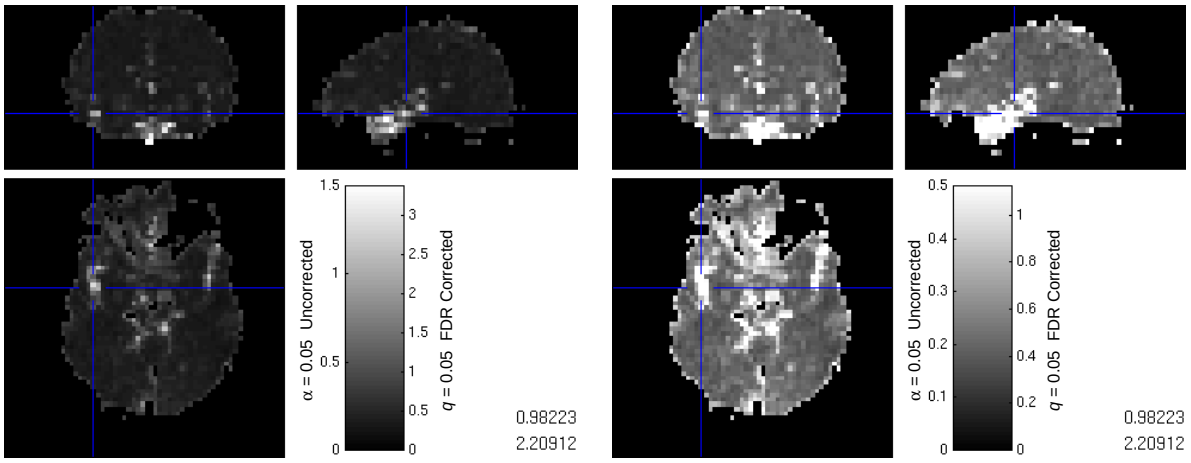
Figure 3 shows the results for the first study; the format is the same as Figure 2. Again, with a glance to the mode-centered gray scale in Figure 3-B we can see that about a 1/2% change is required for a typical voxel to attain corrected significance (about 1/4% uncorrected). A small region of increased variation is indicated with the crosshairs, apparently due to the middle cerebral artery. The impact of this artifact can be readily assessed: To observe a corrected significant change in this region a much larger change is required, from 1.1% to 3.6% (corrected) depending on the voxel. Of course, the most variable voxels are probably centered on the artery and hence not useful indicators of local brain activity. However, if any region surrounding the Sylvian fissure was of importance to this study, this reduced sensitivity would be very important for the interpretation of the results. In particular, the potential for Type II errors in this region should be contemplated and reported.

Figure 4 shows the impact of using a pooled value for the grand mean for the aversive imagery study. The two PCT images, Figure 4-A using voxel-wise baseline, and Figure 4-A using a pooled baseline, are quite similar. Figure 4-C confirms this, showing a scatterplot of voxel-wise baseline PCT versus pooled-baseline PCT; the spread is close to the identity (correlation coefficient 0.87). Since the gray-white contrast is so small in the baseline image, the pooled-baseline PCT appear to be satisfactory approximations for the PCT image. Note from the scatterplot, however, that PCT varies by an order of magnitude; hence there is substantial variation in sensitivity across the brain.



A: PCT B: PCT — Windowed

Figure 2: Assessing artifacts in the second study. A physiologically-related artifact was found in a left temporal-parietal region. Image (A) shows the PCT at that location, and (B) shows the same image windowed such that the modal intensity is approximately centered in the color bar. In the center of the artifact, the PCT is over 5% ($q = 0.05$ FDR corrected; 2.35% uncorrected), and in the region about the artifact, at least a 2.5% change is needed to detect an effect (corrected; 1% uncorrected). In contrast, in the contralateral region a change of less than half of that is detectable. This difference in sensitivity would be of vital importance for a study of laterality effects.



A: PCT B: PCT — Windowed

Figure 3: Assessing vascular variability in the first study. As is common, increased variability was found in the region of middle cerebral artery (MCA). Image (A) shows the PCT at left MCA and (B) shows the same image windowed with a mode-centered colorbar. Note that in most of the brain a change of about 0.6% can be detected (FDR corrected, $q = 0.05$), while in the region of surrounding artifact at least 2.5% change would be needed. In the sagittal view in (B), note how the whole Sylvian fissure shows increased PCT; it is important to characterize the extent of such reduced-sensitivity regions to be aware of potential Type II errors.

(Recall that a pooled-baseline PCT is just the standard deviation images with a scaled colorbar.) The hemifield stimulation exhibited a similar result (correlation coefficient 0.84).

4 Discussion

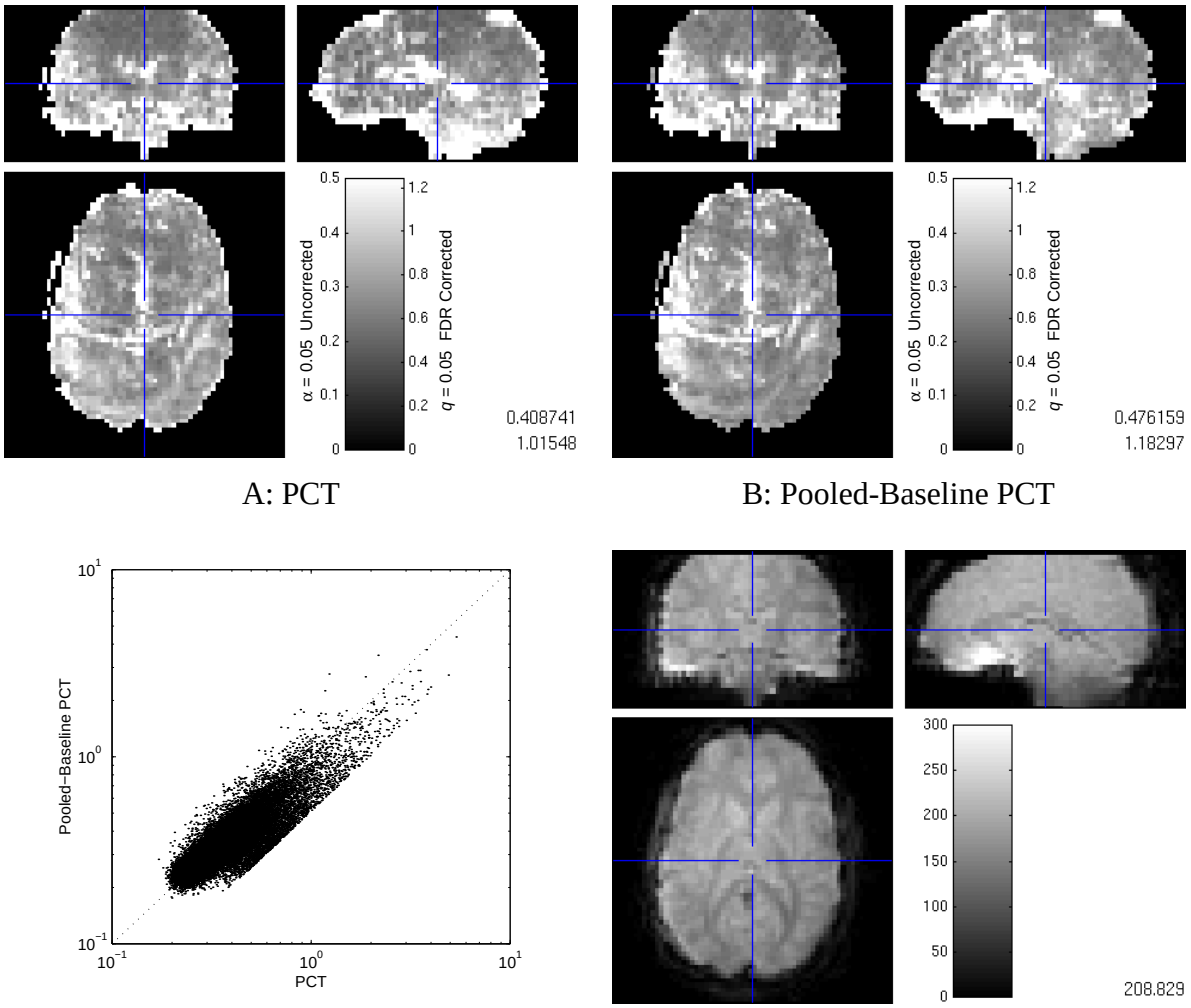
We have described and demonstrated the utility of an image of confidence interval half-length. When viewed with an image of estimated signal, they express exactly the same information as confidence intervals. When viewed alone, it depicts variability in units of the change magnitude required to obtain significance. Thus it conveys both the result of hypothesis tests in intuitive units and the potential for Type II errors. In contrast, a t image can only convey statistical significance and response magnitude in multiples of standard errors.

If images of signal are not quantitative, we recommend visualizing percent change and PCT. Percent change is attractive since it has a concrete interpretation—BOLD effects and nonquantitative PET signals are frequently characterized in percent change. Further, extreme percent change values will readily be identified; in particular, very large percent changes in fMRI indicate the potential for signal from draining veins.

One requirement for creating PCT images is having summary images ($\hat{\delta}$ or $c\hat{\beta}$) that have the same units of the data. We argue that this is a reasonable requirement for all modeling in functional neuroimaging, in that it enriches the interpretation of the summary images. For example, if graymatter has been normalized to 100 and a linear model and contrast preserves units, the contrast image can be interpreted as an approximate percent change image. The success of this approximation, however, depends on the quality of the intensity normalization, which bring us to the next point.

We have found that a simple global mean provides a poor estimate of typical global intensity. Hence we have proposed instead the use of the mode. Our mode estimate uses no topological operations on the image and is quickly computed.

The method is equally applicable to either PET or fMRI. We have shown two fMRI examples where the baseline image is well approximated by a global value. In our use with PET data the heterogeneity of baseline blood flow images was sufficiently large that voxel-wise PCT were preferred.



A: PCT

B: Pooled-Baseline PCT

C: PCT vs. Pooled-Baseline PCT

D: Baseline

Figure 4: Comparison of PCT using local versus pooled baseline. Image (A) shows the standard PCT using a voxel-wise baseline, while image (B) shows the PCT using a pooled baseline; the pooled (i.e. global) baseline value was estimated with the mode of the baseline image (D), here 179.3. Note that (A) and (B) are windowed to better show differences between the images. The images appear very similar, and a log-log scatter plot of the voxel intensities (C) confirm this. This supports the use of a pooled-baseline value when the baseline images are homogeneous.

To the extent that the baseline is uniform across voxels of interest, standard deviation can be directly scaled into PCT. In practice, we routinely scale standard deviation into PCT, even with PET data: We prefer a standard deviation image with an approximate PCTinterpretation to a standard deviation image with arbitrary units.

The ultimate goal of this work was to motivate neuroimaging practitioners to think about the uncertainty of their signal of interest, and how that uncertainty is spatially heterogeneous. By offering a simple means to create interpretable images of signal variability, we hope this has the intended effect.

Acknowledgments

The author gratefully acknowledges Steve Taylor and Tor Wager for use of their data, and Doug Noll for helpful comments.

Appendices

A Estimation of the global brain intensity

This appendix describes the estimation of the mode of brain voxel intensities. We proceed by first estimating a brain-nonbrain threshold, and then estimating the mode of the distribution of brain voxel intensities. While there is an extensive literature on mode estimation using kernel density estimation (see, e.g., Scott, 1992), we simply use a histogram estimate with appropriate bin widths for consistent estimation of mode[1]. Our approach uses no topological operations on the image data and is easily coded and quickly computed.

Estimating Brain-Nonbrain Threshold with the Antimode We estimate a brain-nonbrain threshold using the distribution of all voxel intensities. Our threshold is the location of minimum density between the background and graymatter modes; call this the antimode. Let $f(x)$ be the distribution of intensities in the brain image. Hartigan (1977) shows that a consistent estimator of the antimode is the location of the maximally separated order statistic between modes. Since we don't know the location of modes, we instead just search over the whole density excluding the tails; the tails must be excluded as the global minimum of $f(x)$ will be found there. A crude over estimate of the tails is sufficient, since the antimode estimate will only be perturbed if we include tails with less density than the antimode or exclude the actual location of the antimode. We have found the 10th and 90th percentile to work on all images we have considered. Our threshold estimate is thus

$$T \;=\; \left\{ \frac{1}{2}\left(x_{(k+1)} + x_{(k)}\right) \;:\; k = \underset{0.1n < i < 0.9n}{\text{argmax}} \left(x_{(i+1)} - x_{(i)}\right) \right\} \tag{12}$$

where n is the number of voxels in the image. If k is not unique, we take an average of the locations.

While this works well on continuous-valued image (e.g. a floating point mean image), it does not work with a discrete-valued image (e.g. an integer T2* image). The problem is that the distance between order statistics will be 0 or 1 except at the very extreme tails. Hence if the image is discrete we then revert to a simpler histogram method. We construct a histogram based on all non-tail data

[1] With more and more data, a consistent estimator converges to the true value in probability.

(10th to 90th percentile) and use the location of the minimum bin as the antimode estimate. To construct the histogram we use the bin width rule for the mode (described below). We have found that this serves as a robust estimate of a brain-nonbrain threshold.

Whether through the inter-order-statistic distance or the histogram approach, this antimode estimate is only used to eliminate the lower mode of background voxels and hence does not need to be highly accurate.

Estimating Global Brain Intensity with the Mode Having segmented the brain, we now simply need to estimate the mode of the brain voxel intensities. We use a type of histogram estimate for simplicity and computational efficiency. The optimal histogram bin width for estimating the entire density is order $n^{-1/3}$. However, Scott (1992) shows that using such bin widths to find the mode leads to an inconsistent estimator. Instead, bin widths of order $n^{-1/5}$ are needed to produce a consistent estimate of the mode. For a specific rule, we use bin widths equal to $1.595 \times \text{IQR} n^{-1/5}$ (Scott, 1992, pg 100), where IQR is the interquartile range of the brain voxels. This rule is based on independent Normal data, but has performed quite well on many PET and fMRI datasets. The mode estimate is the location of the maximal histogram bin.

While we do not argue that this mode estimate is optimal in the sense of mean squared error, we have found it to be robust and sufficiently accurate for the purposes of this work. In particular, we have found it more accurate than the simple estimator used in SPM (See Figure 5).

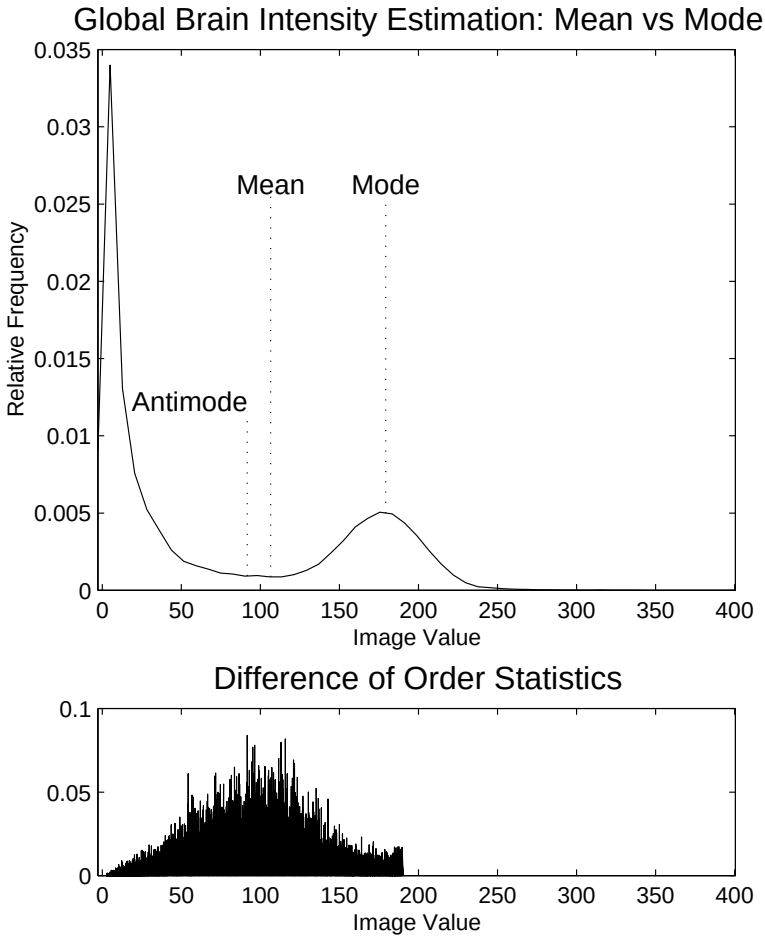


Figure 5: Our mode estimate compared to SPM99's mean estimate. Top plot shows the distribution of intensities of the baseline image shown in figure 4-D. The mode and antimode are as determined by method described in the text. The mean is determined by SPM99, where only voxels greater than 1/8 of global image intensity are considered. The bottom figure shows how the antimode is determined: The difference of order statistics are plotted versus the order statistics; the location of the greatest gap between order statistics is the estimate of the antimode.

References

Genovese, C. R., Lazar, N., & Nichols, T. E. 2001. Thresholding of Statistical Maps in Functional Neuroimaging Using the False Discovery Rate. *NeuroImage, to appear*.

Glover, G.H., & Lai, S. 1998. Self-Navigated Spiral fMRI: Interleaved versus Single-shot. *MRM*, **39**: 361–368.

Graybill, F. A. 1976. *Theory and Application of the Linear Model*. Duxbury Press.

Hartigan, J. A. 1977. Distribution Problems in Clustering. *Pages 45–72 of: Classification and Clustering*. Academic (New York; London).

Holmes, A. P., & Friston, K. J. 1999. Generalisability, Random Effects & Population Inference. *NeuroImage*, **7**(4 (2/3)): S754. Proceedings of Fourth International Conference on Functional Mapping of the Human Brain, June 7-12, 1998, Montreal, Canada.

Neter, J., Wasserman, W., & Kutner, M. 1990. *Applied Linear Statistical Models (3rd Ed.)*. Richard D. Irwin.

Nichols, T. E., & Luo, W. L. 2001. Data Exploration Through Model Diagnosis. *NeuroImage*, **13**(6 (2/2)): S208. Proceedings of Seventh Annual Meeting of the Organization for Human Brain Mapping, June 10-14, 2001, Brighton, England.

Noll, D. C., Stenger, V. A., Vazquez, A. L., & Peltier, S. J. 2000. *Functional MRI*. Springer. Chap. 14, Spiral Scanning in fMRI.

Schervish, M. J. 1995. *Theory of Statistics*. Springer-Verlag.

Scott, D. W. 1992. *Multivariate density estimation. Theory, practice, and visualization*. John Wiley & Sons.