Statistical Parametric Mapping

Geraint Rees

Practical Neurology 2004;4;350-355
doi:10.1111/j.1474-7766.2004.00266.x

HOW TO UNDERSTAND IT

Statistical parametric mapping

Geraint Rees
Wellcome Senior Clinical Fellow, Institute of Cognitive Neuroscience & Institute of Neurology, University College London, 17 Queen Square, London WC1N 3AR;
E-mail: g.rees@fil.ion.ucl.ac.uk
Practical Neurology, 2004, 4, 350–355

INTRODUCTION

Neurologists nowadays regularly encounter statistical parametric mapping in journal articles that report the results of a functional neuroimaging study. The number of such studies has risen dramatically in the last decade, and an understanding of how typical functional MRI (fMRI) experiments are analysed will help the clinician critically reading the literature. Functional imaging studies are typically undertaken to compare brain responses either from a single population in two different conditions (for example, healthy volunteers in the presence vs. the absence of a flickering visual stimulus), or from two populations (for example, neurological patients performing a task vs. control subjects performing the same task).

THE BASIC IDEA

Brain responses are recorded using an imaging technique such as fMRI, which allows the repeated measurement of brain activity at hundreds of thousands of points, or voxels, throughout the brain (Fig. 1). Each voxel represents physiological responses from a small anatomical portion of the brain (typically $3 \times 3 \times 3$ mm). The analysis of such data seeks to identify voxel-by-voxel whether any brain regions show a definite difference in brain activity. As with any other measurement in clinical science, to reliably establish a difference consideration must be given to the *size* of any difference relative to its variability. This requires that a statistical test be carried out to assess any differences in brain activity between the two different conditions, rather than just simply subtracting the activity in the two conditions without any consideration of the variability. The computation and visualization of such tests is known as statistical parametric mapping, and the purpose of this article is to explain how it works.

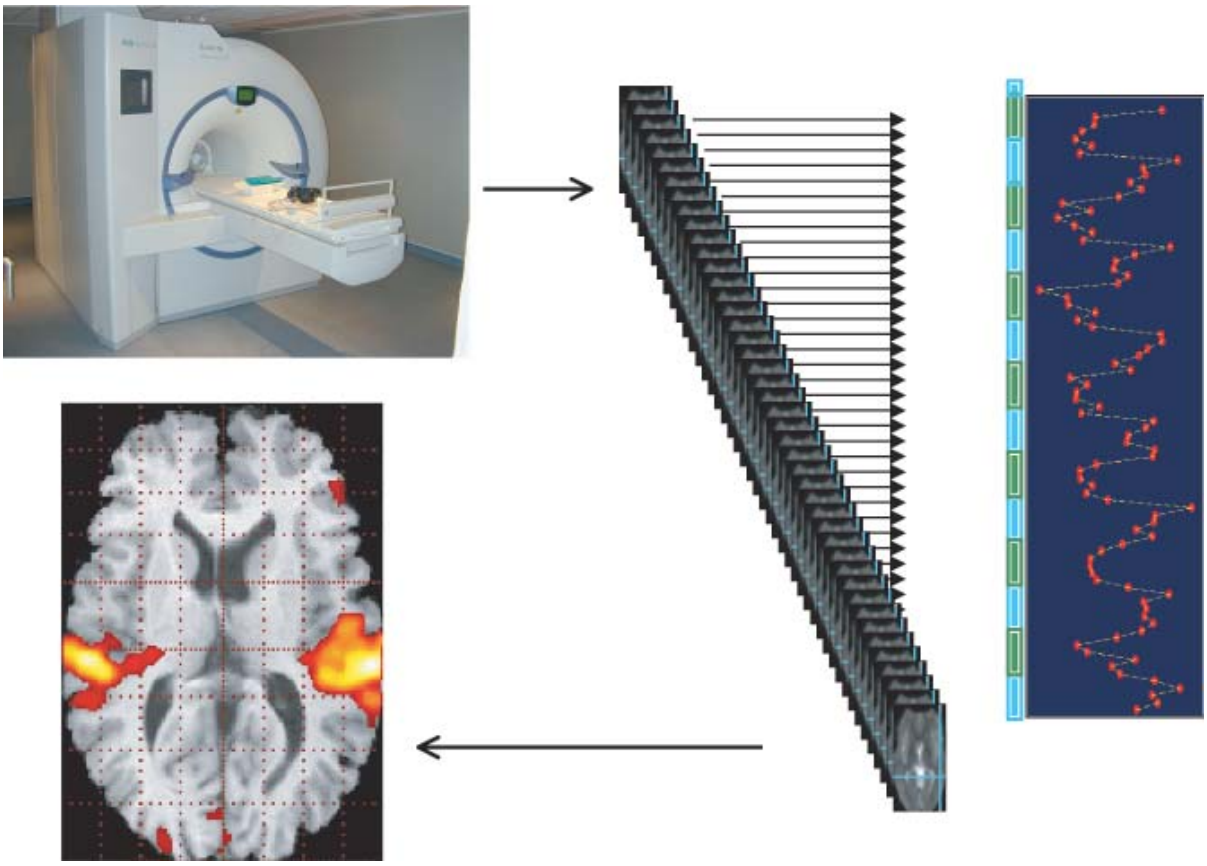

Figure 1 Data analysis. An MRI scanner equipped with echoplanar-imaging capability (in this case a 1.5T Siemens Sonata; top-left) is used to acquire several hundred brain volumes weighted with T2* BOLD contrast. Each slice in the time series has T2* contrast (right) and the responses for one point (indicated by the white cross) are plotted (far right) as a function of time. In this task the subject listened to 30 s epochs of spoken words presented over headphones alternating with 30 s of silence with word presentation indicated by the dark bars in the accompanying key. Fluctuation of the MRI signal correlated with the auditory stimulus presentation is clearly seen (right) and after statistical processing can be displayed as a statistical parametric map (the orange colours corresponding to the level of the statistical parameter) overlaid on a slice of a T1-weighted anatomical volume. Significant bilateral activation in the temporal lobes, centred on primary auditory cortex, is apparent (bottom left).

Statistical parametric mapping is based on a fundamentally simple approach; estimates of brain responses at each and every voxel are analysed using a standard univariate statistical test, and the resulting statistical parameters are then assembled into a three dimensional image, known as the statistical parametric map (SPM). Most commonly, these statistical tests are the same parametric tests already familiar to clinicians, such as simple *t*-tests. In this case, the corresponding SPM is simply the distribution in space of all the *t* statistics from every voxel in the brain. The value at every voxel of such an SPM therefore reflects the value of the Student's *t* for the statistical test at that voxel. Knowing the degrees of freedom, these *t*-values can be looked up in a table in order to convert them into a *P*-value for each voxel for easier interpretation.

Of course, because an SPM often contains hundreds of thousands of different voxels, a certain fraction will come out showing 'significant' activation by chance alone, even if there is no real activation. However, it is possible to adjust the *P*-values to take this multiple comparisons problem into account. One common approach uses a theoretical extension of statistical methodology known as Gaussian Random Field (GRF) theory, the specific details of which are not relevant here. GRF theory provides a method to produce corrected *P*-values that is analogous to the way in which the Bonferroni correction is used for multiple statistical tests on discontinuous data, but now in the context of continuous data (i.e. images of brain responses).

The concepts underlying statistical parametric mapping are straightforward, but in order to clearly understand both the advantages (and

limitations) of the technique, it is helpful to appreciate the basic methodology underpinning the design and analysis of functional imaging experiments. In order to carry out a statistical comparison of brain activity evoked in two different conditions, a number of prior steps are required to extract estimates of brain activity at every voxel. This is because neuroimaging techniques such as fMRI acquire data that are sampled from a continuous process. The time series of data points at each and every voxel must be processed in order to disentangle the effects of the various different experimental conditions, estimating brain responses in each condition, before a statistical test can be performed. This process, and the statistical procedure that follows, will now be described.

FUNCTIONAL MRI TIME SERIES

The most common type of fMRI experiment collects Blood Oxygenation Level Dependent (BOLD) MR image volumes repeatedly in a time series from a set of patients or normal subjects. BOLD images have T2* contrast, which is similar to the T2-images familiar from clinical practice (i.e. CSF appears bright), but is specifically sensitive to changes in the paramagnetic properties of oxy- and de-oxyhaemoglobin. BOLD images reflect the local blood oxygenation, which alters as neural activity changes locally (because synaptic activity is metabolically demanding and so blood flow increases to neurally active regions). Changes in activity between successive BOLD contrast image volumes in a time series therefore can reflect local changes in neuronal activity, but with a lag of a few seconds that reflects the sluggish haemodynamic response to natural activity. The resolution of BOLD contrast MR images is usually lower (perhaps $3 \times 3 \times 3$ mm voxels) than the relatively high resolution T1- and T2-weighted images used in clinical practice, but the acquisition time for each whole brain volume is typically much shorter (2–3 s vs. 10–15 min). Several hundred BOLD contrast volumes are acquired from each subject, creating a time series of several hundred time points at every voxel.

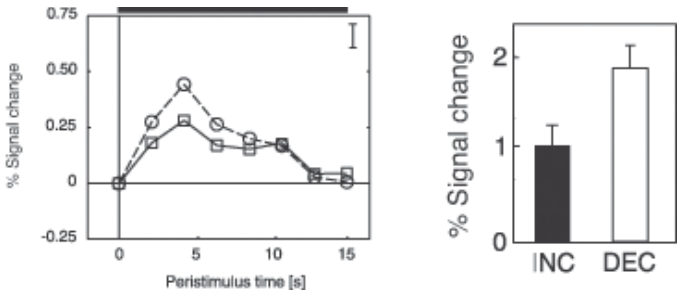

Figure 2 Event-related activation of visual cortex. (Left) Activity in primary visual cortex measured using BOLD contrast functional MRI is plotted as a function of time following either a sustained increase (solid line, squares) or decrease (dotted line, circles) in luminance of a visual stimulus (the horizontal solid bar indicates the duration of the stimulus, the error bar shows ± 1 SE). A delayed haemodynamic response to the neural activity in visual cortex produced by the sudden change is seen, peaking at around 5 s after the change. Note that both decreases and increases in luminance produce a positive response in visual cortex, with decreases evoking a slightly higher signal (Haynes *et al.* 2004). This graph plots raw signal change averaged over many trials, but for statistical comparison an estimate of the response amplitude and variability is required. (Right) Percent signal change modelled by a general linear model assuming a canonical haemodynamic response following each luminance change (black = increment or 'INC' white = decrement or 'DEC', error bars = 1 SE, signal change is expressed as a percentage of voxel mean). The response to decrements is again stronger than that to increments.

EXPERIMENTAL DESIGN FOR FUNCTIONAL MRI

Two types of experimental design are in common use. In a blocked design, the subject is asked to perform a task of interest repeatedly for periods of around 30 s, alternating with one or more other tasks. The aim is to identify the steady-state brain activation during each task epoch, and to identify where in the brain different tasks show different levels of activation (Fig. 2). In an event-related design, individual trials are presented sequentially in a random order, with an interstimulus interval of a few seconds. The aim is to measure the brain activity time-locked to each separate type of trial, and once again to identify where in the brain different types of trial produce different levels of activation. Event-related designs are particularly suited to clinical and experimental situations where the assignment of trial types is made *posthoc* on the basis of a subject's responses. For example, visual extinction following parietal damage does not invariably occur on every trial when a patient is tested with bilateral visual stimulation. The type of trial (extinction or no-extinction) therefore has to be defined posthoc, following the patient's response. Because such trials cannot be identified in advance, the different trial types cannot be presented in a blocked design and so an event-related design must be used.

SPATIAL PREPROCESSING OF NEUROIMAGING DATA

Regardless of the type of experimental design, statistical parametric mapping characterizes brain activity in the same fashion. The international community has now converged on generally accepted standards for data analysis. Several different packages are available (for example, SPM (http:///www.fil.ion.ucl.ac.uk/spm) and FSL (http://www.fmrib.ox.ac.uk/fsl) but all use essentially the same underlying approach with minor variations. Prior to statistical analysis, the time series of several hundred BOLD volumes is typically realigned to the first to correct for head movement during scanning. If several subjects are to be compared, they can then be spatially warped into a standard stereotactic space in order to take account of the different size and shape of brains of different subjects (and thus allow reporting of activated locations in an internationally accepted coordinate system). Sometimes the time series is then spatially blurred, particularly if intersubject comparisons are to be carried out (because even after warping, there can be residual intersubject differences in neuroanatomy that should be disregarded in a group analysis).

STATISTICAL ANALYSIS OF NEUROIMAGING DATA

After spatial preprocessing, statistical analysis can proceed. In order to perform statistical tests on each voxel to identify regions that are significantly activated, estimates of the activity produced by experimental conditions at each voxel must be determined. As there are typically several hundred time-points for every voxel representing the various experimental conditions, a variant of the General Linear Model (GLM) is used to estimate (or model) the brain activity at each voxel in various conditions. A GLM is a hypothesis-driven way of analysing data. It attempts to explain something that has been observed (in this case, the BOLD contrast time series reflecting brain activity measured at each and every voxel) as the weighted sum of a number of hypothesized effects. The hypothesized effects are the neural activity that the experimenter believes has been evoked during a particular trial, convolved (lagged) with the haemodynamic response function (which follows neural activity with a delay of approximately five seconds because of slow neurovascular coupling). For example, if a visual stimulus has been presented for 30 s, the hypothesized neural activity might have the shape of a square wave that is positive (reflecting stimulus-driven neural activity) during stimulus presentation and zero (reflecting the absence of stimulus-driven neural activity) for the remainder of the time series, convolved with the haemodynamic response function.

Having constructed a set of regressors representing the hypothesized brain responses, multiple linear regression is used to determine a set of weights for these regressors at each and every voxel, so that the weighted combination of effects will best correspond to the observed data. These weights represent estimates of brain responses over the MR time-course at each voxel, and will subsequently be used for statistical testing. Regression is performed automatically and independently for every voxel in the imaging time series, yielding a set of parameter estimates (or weights) for each hypothesized effect at every voxel in the brain, and an error term that reflects variability in the observed time series that cannot be accounted for by the hypothesized effects. The size of the parameter estimates, relative to the error term, is then used to calculate an appropriate statistic (e.g. T statistic) for comparisons between conditions of interest, generating (in this case) an SPM{T}. This is what is displayed and reported in pictures and tables in neurological journals as a statistical parametric map.

it is important to remember that what is being displayed is always a *statistical* map, *not* a map of the level of activity per se

DISPLAYING AND READING THE MAPS

SPMs are displayed in different ways, but it is important to remember that what is being displayed is always a *statistical* map, *not* a map of the level of activity per se. Note that the same level of statistical significance can be achieved where the numerical difference between the two things being compared is very small (but the variability in each measurement is also very small) or where the difference is enormous (but the variability is also large). Exactly the same caveat applies to SPMs, which are essentially

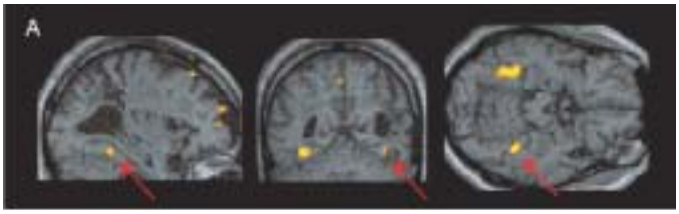

Figure 3 Typical display of a statistical parametric map, thresholded at $P < 0.001$ uncorrected, overlaid on three slices (sagittal, coronal and axial from left to right) of a T1-weighted anatomical image of a patient with parietal extinction following stroke. This SPM represents activity evoked by an extinguished and unseen face stimulus presented in the left visual field, compared to an extinguished house. The arrow indicates activation in the 'fusiform face area' in the right fusiform gyrus, an area known to specifically process faces (and damage to which leads to prosopagnosia). In addition, there is activation in the homologous left 'fusiform face area', as this area is not retinotopically mapped and so responds to faces in either visual field. The presence of such activation indicates that the brain of this patient is capable of distinguishing between faces and other objects presented in the left visual field, even though the patient is unable to consciously report their presence. The full study is published in Rees *et al.* 2002.

very large spatially distributed collections of *t* (or F) values.

Typically, SPM 'activation' maps are either displayed as a three-dimensional rendering onto cortical surface anatomy, or as a two-dimensional overlay onto individual slices of a T1-weighted anatomical image, or perhaps as a Maximum Intensity or 'look-through' Projection (Fig. 3). Usually, SPMs are displayed in the neurological convention (left is on the left) rather than in the radiological convention (left is on the right) (this of course is incredibly confusing for amateurs, which includes most clinical neurologists who are used to looking at brain images in the radiological convention – *Editor*).

When an SPM is displayed, a statistical threshold must be chosen that determines the lower bound of statistical values to display. This procedure is no different from the use of statistics more generally in clinical science, where a threshold (usually $P < 0.05$) must be chosen to assess significance, i.e. that the observed difference has not happened just by chance. In the context of a three-dimensional SPM, the choice of threshold also determines the shape and size of activated brain regions, because only voxels with activations above the chosen threshold are considered 'significant' and therefore displayed. The choice of threshold is not arbitrary, but must reflect an appropriate balance between the risk of false positives (if the threshold is too low) and the risk of false negatives (if the threshold is too high). This is a particular problem for neuroimaging time series because they are made up of image volumes that may contain over 100 000 separate observations (voxels). There is therefore a significant multiple comparisons problem. If the threshold is set too low (for example, $P < 0.05$ uncorrected for multiple comparisons) then many hundreds (or thousands) of false positive voxels may appear in the resultant SPM. On the other hand, if a correction to *P*-values is too stringent, then there is a significant false negative risk. Statistical thresholds must be set to take account of these issues. A threshold of $P < 0.001$, uncorrected, is used for a specific anatomical locus if the experimenter has an *a priori* hypothesis that it will be activated during the experiment. For example, if the experimenter has a strong prior hypothesis that primary visual cortex will be activated by an experimental manipulation, then – based on previous empirical and simulation work – a threshold of $P < 0.001$, uncorrected, is conventionally used (or sometimes an even more stringent value for very specific hypotheses). But for brain regions where the experimenter does not have any *a priori* hypothesis, a more conservative threshold of $P < 0.05$, corrected for multiple comparisons, is conventionally used. This threshold is more conservative because the multiple comparisons correction takes account of the many tens of thousands of voxels (observations) in each brain volume. As mentioned above, the correction is analogous to a Bonferroni correction, but based on GRF theory. Intermediate approaches are also possible, where a correction is applied not to the whole brain volume but to a restricted small volume of interest.

SPMs are displayed in the neurological convention (left is on the left) rather than in the radiological convention (left is on the right)

OTHER APPLICATIONS OF STATISTICAL PARAMETRIC MAPPING

Statistical parametric mapping is a flexible and powerful technique and can be applied to many sorts of imaging data, not just BOLD contrast imaging time series. In fact the technique was first developed in the 1980s in order to analyse Positron Emission Tomography images of brain activation, and is still the standard technique worldwide for this purpose. Similarly, it can be used to analyse images from other modalities, such as SPECT, or other types of MR time series, such as perfusion-weighted and diffusion-weighted sequences. Statistical parametric mapping can also be used to compare T1-weighted anatomical images from patient populations vs. controls, in order to establish whether there are any differences in regional grey matter density. This increasingly popular technique is known as voxel-based morphometry (VBM), and has been used to characterize subtle grey matter changes in brain diseases as diverse as Huntington's chorea, schizophrenia and epilepsy.

As well as permitting the analysis of different types of neuroimaging data, the statistical parametric framework is sufficiently robust to accommodate other types of statistical procedures. Generally the approach using parametric statistics described here is the most powerful (as for clinical science in general), but nonparametric statistical comparisons may be useful in specific situations (for example, they may have greater sensitivity when there are very few subjects in a statistical comparison and so low degrees of freedom). These approaches are all *hypothesis-driven*, in that they require the experimenter to construct a series of regressors representing the hypothesized brain activation. However, *data-driven* approaches to statistical parametric mapping are also available. Such techniques use principal components analysis (PCA) or independent components analysis (ICA) to decompose the imaging time series into a set of time-varying spatially distributed components that can best account for the data. While this can be useful for exploratory analyses, generally the hypothesis-driven approach is most powerful for many experimental situations. Finally, several approaches have emerged recently that allow characterization not just of how activity *within* brain regions changes under different experimental conditions, but how the functional coupling *between* brain regions may change in a similar fashion. One example of such an approach, dynamic causal modelling, represents a potentially powerful way to understand how large-scale networks in the brain interact to produce behaviour.

The most important issue faced by the user of these techniques is the multiple comparisons problem

CONCLUSIONS

In summary, statistical parametric mapping is a flexible and powerful way of testing data acquired using functional imaging techniques to establish patterns of brain activation. Although appearing complicated, it is in fact based on well-established statistical techniques familiar throughout clinical science. For spatially distributed data that form a continuous time series, such as functional neuroimaging data, there are a number of important modifications to allow parametric statistics to be used, but their details are generally not important for a critical appreciation of published data. The most important issue faced by the user of these techniques is the multiple comparisons problem, a consequence of the ability of modern MRI scanners to acquire vast amounts of data at high spatial resolution. However, the power of current analysis techniques can encompass this, while remaining flexible enough to accommodate the next generation of imaging techniques.

ACKNOWLEDGEMENTS

The Wellcome Trust funded this work. I thank Daniel Glaser and Rebecca Roylance for helpful comments.

REFERENCES

Haynes JD, Lotto RB, Rees G (2004) Responses of human visual cortex to uniform surfaces. *Proc Natl Acad Sci U S A*, **101**, 4286–91.

Rees G, Wojciulik E, Clarke K et al. (2002) Neural correlates of conscious and unconscious vision in parietal extinction. *Neurocase*, **8**, 387–93

FURTHER READING

Frackowiak R, Friston K, Frith C *et al.* (2004) *Human Brain Function*, 2nd edition (eds Frackowiak R, Friston K, Frith C, Dolan R, Price C, Zeki S, Ashburner J & Penny W). Academic Press, London.