

A SHORT HISTORY OF STATISTICAL PARAMETRIC MAPPING IN FUNCTIONAL NEUROIMAGING

Prepared as background for an article by Peter Bandettini

The Inception of SPM and Modern-day Brain Mapping

The history of human brain mapping is actually much shorter than one might have anticipated on entering the field. The first brain activation studies harnessed the ability to take images of brain function, successively within the same scanning session. This was facilitated by the use of short half-life radio tracers using Positron Emission Tomography (PET). These techniques became available in the late 80s and the first reports of human brain activation studies appeared in 1988. Up until this time, regional differences among brain scans had been characterised in terms of hand-drawn regions of interest, reducing hundreds of thousands of voxels to a handful of ROI measurements, with a somewhat imprecise anatomical validity. The idea of making voxel-specific statistical inferences, through the use of statistical parametric maps, emerged from the then young brain mapping community in response to the clear need to make valid inferences about brain responses without knowing where those responses would be expressed. Statistical parametric maps refer to image processes of a statistical parameter. This statistic (usually the T or F statistic) is carefully chosen to relate directly to the effect one is interested in. The ensuing SPM can be thought of as an X-ray of the effects significance. In the absence of any effect the distribution of the statistic conforms to its null distribution thereby enabling classical inferences (through rejection of the null hypothesis) in declaring a regionally specific activation. The first SPM was used to establish functional specialization for colour processing in 1989 (Lueck et al 1989). The methodology and conceptual underpinning of statistical parametric mapping were described in two papers in 1990 and 1991. The first paper (Friston et al 1990) entitled "The Relationship Between Global and Local Changes in PET Scans" presented the formal basis of statistical parametric mapping using both the T & F statistics. This may seem an odd title to introduce statistical parametric mapping but it is interesting to understand its motivation from a historical perspective. Previously, average values from ROIs had been entered into multi-factorial analysis of variances. This approach had become established in the analysis of auto-radiographic data in basic neuroscience and glucose uptake scans in human subjects. The critical thing here was that region was treated as a level of a factor. This meant

that the regional specificity of a particular diagnosis or treatment was reflected in the region by treatment interaction. In other words a main effect of treatment per se was not sufficient to infer a regionally specific response. This is because some treatments induced a global effect that was expressed in all ROIs. The issue of how to deal with global effects in SPM was, therefore, one of the first major conceptual issues in its development. The approach taken was to enter global activity as a confounding variable in analysis of covariance thereby imbuing detected changes with a regional specificity that could not be explained by global changes. The whole issue of regional versus global changes and the validity of global estimators was debated for several years with many publications in the specialist literature. Interestingly, it is a theme that enjoyed a reprise with the advent of fMRI (e.g. Zarahn et al) and still attracts some research interest today.

The second paper (Friston et al 1991) was entitled “Comparing Functional Images: The Assessment of Significant Change”. This was a critical paper that framed a fundamental problem that is unique to inference in human brain mapping. Namely, the need to account for the enormous number of statistical tests used in the mass univariate approach of SPM. Clearly performing a statistical test at each and every voxel engendered an enormous false-positive rate using conventional and unadjusted thresholds to declare an activation was significant. The problem was further compounded by the fact that the data were not spatially independent and a simple Bonferroni correction was inappropriate. This was the second major conceptual theme that occupied many of the theorists trying to characterise functional neuroimaging data. What was needed was a way of predicting the probabilistic behaviour of SPMs, under the null hypothesis of no activation that properly accounted for the smoothness or spatial correlations among nearby voxels. Distributional approximations were derived using the theory of stochastic processes and estimates of the smoothness. It quickly became evident that the early heuristic proofs provided in Friston et al (1991) were closely related to almost identical results in the theory of Gaussian fields. Gaussian fields are stochastic processes that conform very nicely to realizations of brain scans under normal situations. Within a year the technology to compute corrected P values was embedded in Gaussian field theory (Worsley et al 1992). Although the basic principals were established at this time there have been many exciting mathematical developments with extensions to different sorts of SPMs and the ability to adjust the P values for small volumes of interest (see Worsley et al 1996). Robert Adler, one of the worlds contemporary experts in Gaussian Field Theory, who had abandoned it 10 years before it was embraced by the imaging community, has now returned to GRF with its renaissance in brain imaging.

The name Statistical Parametric Mapping was chosen carefully for a number of reasons. Firstly, it represented a nod to the acronym of significance probability mapping developed in the field of EEG. Significance probability mapping involved creating interpreted pseudomaps of P values to render the spatio-temporal organisation of evoked electrical responses more apparent on inspection. The second reason was somewhat more historical: In the early days of PET many images were derived from the raw data reflecting a number of different physiological parameters (e.g. oxygen metabolism, oxygen extraction fraction, regional cerebral bloodflow etc). These were referred to as parametric maps. In exactly the same way as these physiological parameters were non-linear functions of the original data a statistical parametric map is, likewise, a non-linear function of the original data. The distinctive thing about SPMs is that they have a known distribution under the null hypothesis. This ensues from the fact that they are predicated on a statistical model of the data that has an explicit error term. In essence by dividing an activation effect by its standard error, to produce a T statistic one, is normalising the differences between two brain scans and a voxel specific fashion. This normalisation causes the statistic to be identically behaved, everywhere in the context that there is no activation. One important controversy among the inceptors of statistical parametric mapping was whether that variability or error variance was the same from brain region to brain region. The Friston group maintained that it was not and have adhered to voxel-specific estimates of error. The Worsley group considered that the differences in variability could be disregarded. This allowed them to pool their estimator over voxels to give very robust and sensitive SPMs but at the expense of questionable validity. Interestingly, this issue has not dogged fMRI where it is generally accepted that error variance can change from voxel to voxel. The third motivation for the name statistical parametric mapping was that it reminded people one is dealing with parametric statistics that assume the errors are additive and Gaussian. This is in contra-distinction to non-parametric approaches that are generally less sensitive, more computationally intensive but do not make any assumptions about the distribution of error terms. Although there are some important applications of non-parametric approaches they have not been widely adopted by the imaging community. This is largely because brain imaging data conform almost exactly to parametric assumptions by the nature of image reconstruction and post processing that is applied to the data.

The PET Years

In the first few years of 1990s many landmark papers were published using PET and the agenda for many functional neuroimaging programs was established. SPM became established as the

most popular and objective way to characterise brain activation data. It was encoded in Matlab and used extensively by the MRC Cyclotron Unit at the Hammersmith Hospital in the UK and was then distributed to collaborators and other interested units around the world. The first people outside the Hammersmith group to use SPM were researchers at NIMH (Jim Haxby and Leslie Ungerleider). Within a couple of years SPM became the community standard for analyzing PET activation studies and the utility of SPMs was largely taken for granted. By this stage, SPM had become synonymous with the union of the general linear model and Gaussian field theory. Although originally framed in terms of AnCova it was quickly realized that any general linear model could be harnessed to produce a SPM. This spawned a simple taxonomy of experimental designs and their associated statistical models. These were summarized in Friston et al (1995a) in terms of subtraction or categorical designs, parametric designs and factorial designs. The adoption of factorial designs was one of the most important advances in this era. The first study to employ a factorial design focussed on adaptation during motor learning and was quickly followed by studies looking at the interaction between a psychological and pharmacological challenge in psychopharmacological studies (Friston et al 1992 a,b). The ability to look at the effect of changes in the level of one factor on activations induced by another facilitated a complete rethink of cognitive subtraction and pure insertion and the awareness of context sensitive activations in the brain. The enormous latitude afforded by factorial designs is reflected in the fact that nearly every study in the modern day literature is multi-factorial in nature. Because SPM used a completely general framework, factorial designs were trivial to analyse and could be easily specified by the appropriate contrast of parameter estimates to produce a corresponding T or F statistic.

Although strictly speaking, not a part of SPM, there was another important development in the early 90s that made the development of fMRI very welcome. This was a growing appreciation that functional specialization and regionally specific activations were not the complete story in discerning functional brain architectures. There was an increasing interest in functional integration and understanding the discourse among different brain areas during the execution of a sensori motor or cognitive process. This was reflected in attempts to take the characterisation of functional imaging time series beyond the mass univariate approach to explicit multi-variate approaches that could look at the coupling among brain areas. The first eigenimages, based on principal component analyses of PET time-series appeared in 1993 (Friston et al 1993). Despite the simplicity and data-led elegance of these characterizations they have not caught on, or challenged conventional SPM-like analyses. This reflects the fact that the scientific process in neuroimaging is essentially Popperian and is almost universally, driven by hypotheses. This approach calls on classical inference and is exactly the need met by Statistical

Parametric Mapping. However, the interest in functional connectivity gave rise to the notion of effective connectivity (the influence that one neural system exerts over another); A notion that was championed by the pioneering work of Randy Mackintosh using structural equation modeling.

The advent of fMRI

In 1992 at the annual meeting of the Society of Cerebral Blood Flow and Metabolism in Miami, Florida, Jack Belliveau presented, in the first presentation of the opening session, his provisional results using photic stimulation with fMRI. This was quite a shock to the imaging community that was just starting to enjoy a sense of consolidation and complacency. Most of the outstanding problems had been resolved, community standards had been established and the summer years of imaging neuroscience lay ahead for effortless indulgence. It was immediately apparent that this new technology was going to radically re-shape brain mapping, the community was going to enlarge considerably and that those researchers already within it were going to have to re-skill themselves. The enormous benefits of fMRI were clear, in terms of the ability to take many hundreds of scans within one scanning session and to repeat these sessions indefinitely in the same subject. Some people say that the main advances in a field, following a technological breakthrough, are made within the first five years. Imaging neuroscience must be unique in the biological sciences in that almost exactly at the end of the first five years following the inception of PET activation studies, fMRI arrived. The advent of fMRI brought with it a new wave of innovation and enthusiasm which carried imaging science, along with genomics and molecular biology to the forefront of scientific endeavor at the end of the last millennium.

From the point of view of SPM there were two problems, one easy and one hard. The first problem was how to model evoked hemodynamic responses in fMRI time series. This was an easy problem to resolve by virtue of the fact that SPM could embrace any general linear model, including first order (ie linear) approximations to the way hemodynamic responses were caused. The general linear model here was simply a linear time invariant system or convolution model that was detailed explicitly in Friston et al (1994). The only remaining issue was the form of the convolution kernel or hemodynamic response function that should be adopted. Stimulus functions encoding the occurrence of a particular event or experimental state (e.g. boxcar functions) were simply convolved with the HRF to form regressors in a general linear model (c.f. multiple linear regression). The P values associated with the ensuing T statistic are mathematically identical to those based upon the corresponding correlation co-efficient.

Around this time the pioneering work of Peter Bandettini has rendered the use of the correlation co-efficient a very popular device. The formal equivalence between SPM regression approaches and correlation co-efficient based approaches took some time to percolate into the popular culture.

The second problem that SPM had to contend with was the fact that successive scans in a fMRI time series were not independent. In PET each observation was statistically independent of its precedent but in fMRI colored noise in the time series rendered this assumption invalid. The very existence of these temporal correlations originally met with some skepticism but are now established as an important aspect of fMRI time series. The SPM community tried a series of heuristic solutions until it arrived at the formulas presented in Worsley & Friston (1995). These provided a completely general framework that retained its exact connection with earlier techniques but embodied any arbitrary form for serial correlations among the error terms. This was based on the Satterthwaite conjecture and is formally identical to the non-specificity correction developed by Giesser and Greenhouse in conventional parametric statistics. The issue of serial correlations and more generally non-sphericity is still important and attracts much research interest particularly in the context of the best temporal filtering to be applied to fMRI time series.

New workers in the fMRI community readily adopted many of the developments and lessons learnt from the early days of PET. Among these were the use of the standard anatomical space provided by the atlas of Talairach and Tournoux (1988) and many conceptual issues relating to experimental design and interpretation. Many debates that had dogged early PET research were rapidly resolved in fMRI, for example, "What constitutes a baseline"? This question, which preoccupied the whole community at the start of PET, appeared to almost be a non-issue in fMRI with the use of elegant and well controlled experimental paradigms. Other issues such as global normalisation were briefly revisited given the very different nature of global effects in fMRI (multiplicative) relative to PET (additive). One issue though remained largely ignored by the fMRI community and is still only now being addressed through peer review and consensus. This is the issue of correcting the P values for the multiplicity of tests performed. While people using SPM quite happily adjusted their P values using Gaussian field theory, others seemed to discount the problem as unimportant. There appeared in the literature a number of reports that used uncorrected P values, a tendency which still confounds many editorial decisions today. It is interesting to contrast this in a historical perspective with the appearance of the first PET studies. When people first started reporting PET experiments there was an enormous concern about the rigor and validity of the inferences that were being made. Much of this concern came

from outside the imaging community who, understandably, wanted to be convinced that the hotspots that they saw in papers reflected true activations as opposed to noise. The culture at that time was hostile to the capricious reporting of potentially specious results and there was a clear message from the broader scientific community that the issue of false-positives had to be resolved. This was the primary motivation for developing the machinery to adjust P values to protect against family wise false-positives over the entire brain (ie Gaussian field theory). In a sense SPM was a reaction to the very clear mandate set by the larger community, to develop a valid and rigorous framework for making inferences about activation studies. Once this had been achieved the agenda disappeared. By the time fMRI arrived it was able to enjoy an acceptance that was much less critical. Only within the past year or two have editors and peers been insisting that the P values be adjusted for the volume of SPM searched. Perhaps, one explanation for this retrograde step was that the equations underlying random field theory-based adjustments are quite intimidating and that this area of research was already fairly deep and mature when fMRI started. This may have precluded a graceful and easy transfer of this important know-how to new fMRI units. With the increasing use of SPM for fMRI (there are now over 1,000 active discussants on the SPM Email List), the distinction between PET and fMRI cultures has now almost disappeared.

In the mid 90s there was lots of research with fMRI some of it was novel and inspirational, other programs consolidated and refined earlier findings with PET. From a methodological point of view notable advances included the development of event-related paradigms that allowed experimenters to escape the constraints on cognitive set imposed by block designs and, secondly, the use of retinotopic mapping to establish the organisation of cortical areas in human visual cortex. This incidentally inspired a whole sub-field of brain flattening and cortical surface mapping that is an important endeavor in early sensory neuroimaging. From the point of view of SPM there were three important challenges to be addressed. The first involved a refinement of the models of evoked responses. The convolution model in the context of a linear time invariant system had become a cornerstone for fMRI with SPM. In 1995 a device was described in which evoked responses were modeled as a linear combination of temporal basis functions (Friston et al 1995b). This was important because it allowed one to define not a single HRF but a family of HRFs that could differ voxel by voxel. This family was determined by the space spanned by the temporal basis set used. This general approach, using temporal basis functions found an important application in the analysis of event-related fMRI data. The general acceptance of the convolution model was consolidated by the influential paper of Boynton year later (Boynton et al 1996). However, at this time people were starting to notice some non-linearities in fMRI responses that were formulated, in the context of SPM, as a

Volterra series expansion of the stimulus function. This again was simple because the Volterra series is another linear model (c.f. a Taylor expansion).

The second issue, that concerned the developers of SPM, was borne of the growing corpus of event related fMRI studies. This was the efficiency with which responses could be detected and estimated. Using an analytical formulation, it was simple to show that the boxcar paradigms were in fact much more efficient than event-related paradigms, but event related paradigms could be made efficient by randomizing the occurrence of particular event such that they “bunched” together to increase experimental variance. This was an interesting time in the development of data analysis techniques because it enforced a signal processing perspective on the general linear models employed. There is still ongoing debate as to whether it is important to maximize efficiency for detecting an activation, as opposed to estimating the form of evoked response.

The third area motivating the development of SPM is unique to fMRI and reflects the fact that many scans can be obtained in many individuals. Unlike in PET, the within-subject scan-to-scan variability can be very different from the between-subject variability. This difference in variability is referred to as non-specificity and generally calls for random as opposed to fixed-effects analyses. This speaks to the whole notion of linear hierarchical observation models for fMRI data and the interesting possibility of parametric empirical Bayesian treatments of these models. Because SPM only had the machinery to do single level (fixed effects) analyses a device was required to implement mixed or random-effects analyses. This turned out to be relatively easy and intuitively accessible. Subject-specific effects were estimated in a first-level analysis and the contrasts of parameter estimates (e.g. activations) were then re-entered into a second level SPM analysis. This recursive use of a single-level statistical model is fortuitously equivalent to multi-level hierarchical analyses (assuming that the designs are balanced and there are no missing data).

A Retrospective

What has PET and fMRI told us over the past decade? One could recount numerous examples from imaging neuroscience where our understanding of functional brain architectures has been substantially and explicitly informed by brain mapping studies. However, the main impact of functional neuroimaging has not been a paradigm shift per se but a translation of extant conjectures into empirically established and useful dogma. Perhaps the clearest example is the

notion of functional segregation and specialization in the brain. This was a conjecture based upon almost a century of lesion and stimulation experiments, supplemented more recently with detailed electrophysiology and neuroanatomy in basic neurosciences. However, before 1988 there was no way of establishing whether a particular cortical area was indeed specialised for a particular function or that that function was segregated anatomically to a particular cortical region. Only with the first neuroimaging evidence for things like colour and motion specific processing did the notion of functional specialization become fact. This was not a revolution but had a subtle and fundamental impact on our neuroscientific understanding. Alternative ideas about how the brain works, or is organised, disappeared gently from discussion allowing people to focus on the more biologically plausible mechanisms of perceptual synthesis and cognition. The prominence of imaging neuroscience in cognitive neuroscience and all its related meetings testifies to the central role that functional brain mapping has now attained in the neurosciences. It also speaks to the demise of cognitive science in relation to cognitive neuroscience. The boxological and abstract constraints of cognitive scientists are not, in themselves, invalidated by neuroimaging. Indeed, empirical brain results are somewhat orthogonal to cognitive science in the first place anyway. However, the interesting questions now relate to the validity of cognitive neuroscience models of processing or the principles that underlie early visual processing insofar as they fit comfortably with empirical neuroimaging results. In short, the question has changed from “How can it be done?” to “How does the brain do it?” Neuroimaging has fundamentally re-framed most aspects of neuroscience and in particular cognitive neuroscience. Young researchers today have little interest in learning the skills required to do single or multi-unit electrode recordings in awake behaving primates but are eager to learn how to study humans with fMRI. This is a cultural brain-drain that some find sad but is inevitable given the vastly accelerated programs of research afforded by non-invasive functional imaging over conventional electrophysiological techniques in basic neuroscience. A similar reorientation is evident in psychology. In the early years functional neuroimaging was undertaken by small groups of experts who happen to be in the right place at the right time and have the right eclectic mix of expertise. As a methodological aspects have become more established, and skills more easily disseminated many functional neuroimaging programs have been taken over by psychologists who would now prefer themselves to be thought of as cognitive neuroscientists. The ability to measure the brain responses at 100,000 locations is much more attractive than measuring a single variable such as perceptual threshold or reaction time. This is particular evident in neuropsychology where the lesion deficit model is undergoing a complete re-evaluation. The concept of necessary and sufficient brain systems only has meaning in the context of functional neuroimaging which can establish the sufficiency of a particular system to support a given task. Prior to the 1990s only those brain systems

necessary for a particular task could be identified through patient studies. Only now is it emerging that the brain is organised as a series of degenerate brain systems where a particular function can be met by any number of subs-sets of cortical areas. This will have profound implications for neuropsychological inference based purely upon lesions and will rewrite our understanding of cognitive deficits following brain damage. Similar arguments can be made in relation to the interactions between brain regions. These are now available for empirical study as opposed to the abstract and unconstrained lesions made to mathematical models of parallel distributed processing popular in the 1980s. In summary, functional neuroimaging has had an impact on almost every neuroscience field in a way which has been constructive because it draws people in, and enables them to address the questions posed in their disparate fields.

Acknowledgements: Thank you to Marcia Bennett for editorial assistance and the Wellcome Trust for supporting this work

References

- Lueck CJ Zeki S Friston KJ Deiber NO Cope P Cunningham VJ Lammertsma AA Kennard C and Frackowiak RSJ The colour centre in the cerebral cortex of man *Nature* 1989; 340:386-389
- Friston KJ Frith CD Liddle PF Dolan RJ Lammertsma AA and Frackowiak RSJ The relationship between global and local changes in PET scans *J Cereb Blood Flow Metab* 1990; 10:458-466
- Friston KJ Frith CD Liddle PF and Frackowiak RSJ Comparing functional (PET) images: The assessment of significant change *J Cereb Blood Flow Metab* 1991; 11:690-699
- Friston KJ Frith C Passingham RE Liddle P and Frackowiak RSJ Motor Practice and neurophysiological adaptation in the cerebellum: a positron tomography study *Proc R Soc. Lond Series B* 1992a; 248:223-228
- Friston KJ Grasby P Bench CJ Frith C Cowen P Little P Frackowiak RSJ and Dolan R Measuring the neuromodulatory effects of drugs in man with positron tomography *Neuroscience Letters* 1992s; 141:106-110
- Friston KJ Frith C Liddle P and Frackowiak RSJ Functional Connectivity: The principal component analysis of large data sets *J Cereb Blood Flow Metab* 1993; 13:5-14
- Friston KJ Jezzard PJ and Turner R Analysis of functional MRI time-series *Human Brain Mapping* 1994; 1:153-171
- Friston KJ Holmes AP Worsley KJ Poline JB Frith CD and Frackowiak RSJ Statistical Parametric Maps in functional imaging: A general linear approach *Human Brain Mapping* 1995a; 2:189-210
- Friston KJ Frith CD Turner R and Frackowiak RSJ Characterizing evoked hemodynamics with fMRI *NeuroImage* 1995b; 2:157-165
- Worsley KJ and Friston KJ Analysis of fMRI time-series revisited - again *NeuroImage* 1995; 2:173-181
- Worsley KJ Marrett S Neelin P Vandal AC Friston KJ Evans AC A unified statistical approach for determining significant signals in images of cerebral activation *Human Brain Mapping* 1996; 4:58-73