Statistica Sinica **18**(2008), 1203-1232

Editorials

Statistical Relativism in Neuroimaging

This editorial considers advances in the analysis of imaging data over the past decade, with a particular emphasis on the role of generative or forward models[1]. It highlights the distinction between establishing statistical dependencies among variables and the identification or inversion of models that generate data. This relates closely to the distinction between exploratory and hypothesis-led analyses. We will pay special attention to the articles in this special issue, when motivating and illustrating trends in statistical approaches. To make this review entertaining (and a little contentious), we invoke the notion of *statistical relativism*. Linguistic relativism refers to the reification of concepts entailed by the language used to describe them. In a similar sense, we will use statistical relativism to mean a reification of models used to disclose statistical dependencies as generative models of data. In some instances, this reification obscures the fundamental difference between true causal models and probabilistic mappings. We will review recent developments in light of this distinction and argue in favour of generative models in imaging neuroscience.

1. Introduction

This special issue brings together many exciting developments that exemplify advances made over the past years in characterising neuroimaging data and, in particular, neurophysiological time series that can be measured with brain imaging. We will focus on functional magnetic response imaging (fMRI) data analysis but many of the issues also attend the analysis of electroencephalographic data and structural images. The basic premise of this editorial is that these advances fall into two classes. On the one hand, there are developments that rest on informed and biophysically constrained generative models, while on the other we have analyses based upon tests for statistical dependencies between experimental and measured variables or among subsets of measured data. This distinction emerges in many domains of data analysis; ranging from the detection of regionally specific activations, through to sophisticated

[1] The Wellcome Trust funded this work. We would like to thank Marcia Bennett for preparing this manuscript and Justin Chumbley for highlighting the literature on linguistic relativism.

analyses of functional integration and distributed brain processing. In what follows, we will clarify the distinction between generative models and models used to test for dependencies and then review some of the major trends in statistical neuroimaging, with this distinction in mind.

2. Generative Models

The distinction between generative models and models of dependency dates back to the first analyses of fMRI time series. At the inception of fMRI, two distinct approaches to data analysis emerged. One used linear convolution models of hemodynamic responses. Here, stimulus functions encoding experimentally evoked neuronal activity were convolved with a hemodynamic response function to provide explanatory variables or regressors for observed fMRI responses (Friston, Jezzard and Turner (1994)). The other approach simply looked for correlations between delayed stimulus functions and the observed response (Bandettini, Jesmanowicz, Wong and Hyde (1993)). Mathematically, if the delay is implemented by convolution with a hemodynamic response function, then the inferences furnished by both approaches are identical; because both are based upon the same general linear model. However, there is a conceptual difference between the linear convolution model, which can be used to generate data, given the parameters of the hemodynamic response function (HRF) and the correlation analysis, which simply establishes a statistical dependency between explanatory and response variables. Both significant [non-zero] HRF amplitudes and significant correlations are sufficient to establish a regionally specific activation in the brain. However, the correlation coefficient approach can do no more than this. In contrast, generative models based upon convolution models continued to develop; the first development was in terms of nonlinear or generalised convolution models. Subsequently, state-space-models based upon the biophysics of the hemodynamic response were introduced in the context of Dynamic Causal Modelling (DCM).

The hemodynamic response function remains a key focus of research in statistical neuroimaging; for example, Loh, Lindquist and Wager (this issue) present an analysis of model residuals that "could be a valuable tool in assessing violations of statistical assumptions and informing about differences in the shape and timing of the HRF across the brain." The key point here is that generative convolution models became increasingly constrained by the biophysics and physiology of evoked hemodynamic responses, which enabled inferences about the underlying dynamics and mechanisms generating data. In general terms, generative models are statistical models that act as

observation models with an explicit generative process. As noted by Guha and Biswas (this issue), "constructing models for neuroscience data is a challenging task, more so when the data sets are of [a] hybrid nature. The models have to be physiologically meaningful, as well as statistically justifiable." In the context of static data, Zhou and Wang (this issue) present a nice example of surface shape-analysis. These authors present a novel "statistical surface analysis framework that aims to accurately and efficiently localize regionally specific shape changes between groups of 3D surfaces." Invoking *shape* as an explicit cause of observed image data is perhaps one of the simplest and most fundamental examples of generative modelling, an example that is at the heart of modern theories of vision (e.g., Marr (1976)). See also Chung, Hartley, Dalton and Davidson (this issue) who model the cortical surface using spherical harmonics.

The identification or inversion of generative models, given some data, enables conditional inferences about the parameters of these models, and critically, comparison of different generative models (Penny, Stephan, Mechelli and Friston (2004)). This model comparison has been an important theme in recent developments, because it embodies the scientific process, in terms of hypothesis testing. This is because hypotheses can be framed in terms of competing generative models of the same data and model comparison can be used to adjudicate among them. A nice example of this is provided in Guha and Biswas (this issue) who compare hidden Markov and state-space models of local field potentials and neuronal firing using "standard model selection criteria like AIC and BIC". In contrast, techniques that rely solely on establishing statistical dependencies or probabilistic mapping between behavioural and physiological data, or between physiological data acquired from different parts of the brain, provide no machinery for model comparison beyond the existence of that mapping. In what follows, we look at three examples of generative modelling and then consider three examples of models that aim to detect statistical dependencies.

3. Examples of Generative Models

3.1 Effective connectivity and models of distributed brain responses

Perhaps the best example of generative modelling is the elaboration of dynamic models of multivariate responses entailed by distributed processing in the brain. This ambition is illustrated nicely by the work reported in Sánchez-Bornot, Martínez-Montes, Lage-Castellanos, Vega-Hernández and Valdés-Sosa (this issue). These

authors try to estimate voxel-based effective connectivity. Effective connectivity corresponds to coupling parameters in models of distributed brain responses. Efficient identification of these models is very difficult because there are many more parameters than observations. The authors harness the known sparsity of brain connections (through the use of Local Quadratic Approximation and the Minimisation-Maximisation algorithm) to access the unknown coupling parameters. This is a nice example of using the known architectural attributes of systems generating data to finesse statistical models of those data.

Another important example of generative modelling is Dynamic Causal Modelling (DCM; Friston, Harrison and Penny (2003)). This is the direction on which my group has focused; a direction which we see as a fairly simple generalisation of the early convolution models for fMRI. In other words, the convolution models used in statistical parametric mapping are formally identical but special cases of Dynamic Causal Models of distributed responses. In DCM, one regards the data as being caused by perturbations of hidden neuronal states by experimental inputs. These perturbations produce neuronal dynamics, through neuronal interactions and are passed through static nonlinear functions to form observed responses. Inversion of these models furnishes conditional densities (e.g., conditional mean and precision) on the parameters of the underlying neuronal and hemodynamic models and their marginal likelihood for model comparison. It is interesting to note that DCM was developed to infer on effective connectivity between nodes in distributed brain networks exhibiting evoked responses. The definition of effective connectivity calls on generative models, in the sense that effective connectivity is defined as the causal influence that one neuronal system exerts over another. Contrast this with the definition of functional connectivity, which is usually taken to mean statistical dependencies between time-series acquired from different parts of the brain. Functional connectivity is generally assessed with correlations between remote regional time series or, in more temporarily resolved electroencephalography data, coherence analyses with frequency specificity (see Ombao, Shao, Rykhlevskaia, Fabiani and Gratton; this issue). We will return to advances in functional connectivity later.

3.2 Models of electromagnetic sources

Perhaps the most obvious examples of generative or forward models in neuroimaging are those used to model observed channel data in electroencephalography (EEG) and magneto-encephalography (MEG), in terms of source activity in the

brain. There have been some remarkable developments in this field that rest largely upon the Bayesian inversion of constrained forward models of electromagnetic sources. These generative models embed electromagnetic forward models based upon classical electrodynamics. However, they also include hierarchical constraints on the deployment of sources, which regularise the ill-posed inversion problem. The main advances in this area pertain to the empirical optimisation of spatial priors on where these sources are, such as smoothness or sparsity constraints. This speaks to an important and generic development in generative models, namely the use of hierarchical models to explain data. Without exception, models with a hierarchical form induce empirical priors, which bring important benefits in terms of increased precision or accuracy on the estimated parameters, such as source location and activity. An excellent example of this is the approach described by Vega-Hernández, Martínez-Montes, Sánchez-Bornot, Lage-Castellanos and Valdés-Sosa (this issue). These authors present a general formulation of the inverse problem "as a Multiple Penalized Least Squares model, which encompasses many of the previously known methods as particular cases (e.g., Minimum Norm, LORETA)." Furthermore, as the authors note, "new types of inverse solutions arise since recent advances in the field of penalized regression have made [it] possible to deal with non-convex penalty functions, which provide sparse solutions."

3.3 Multimodal or fusion models

Another compelling example of generative models in neuroimaging are models that try to explain different sorts of data. A very nice example can be found in Guha and Biswas (this issue). These authors introduce various techniques for inverting models of multimodal (hybrid) time series data. As an example, they consider "local field potentials (which is a continuous time series) and nerve cell firings (which is a point process) of anesthetized mice". Perhaps the best known example of these *fusion* models are generative models that are framed in terms of neuronal activity that can generate or explain observed electrophysiological responses in EEG or MEG and, at the same time, explain hemodynamic responses observed with fMRI (e.g., Daunizeau, Grova, Marrelec, Mattout, Jbabdi, Pélégrini-Issac, Lina and, Benali (2007)) Fusion models are still at an early stage of development but represent, for some, the holy grail of generative models in neuroimaging. The advantage of these models is that they harvest complimentary spatial and temporal constraints from fMRI and electrophysiological measurements, respectively.

In the next section, we turn to models that are used to establish dependencies between sets of variables. In these models the form (and direction) of the mapping from one set of variables to another is really incidental to the goal of establishing that the mapping exists. However, there is an understandable temptation to think of the generative process, behind the data, in terms of the model used to test for a mapping. We will discuss this conceptual conflation of generative models and probabilistic mappings in terms of linguistic relativism and highlight some of its consequences.

4. Examples of Inference on Mappings

Linguistic relativism began with a question posed by Franz Boas in the 1900s, with his work with the Inuit tribe. Could a culture's evolution be described with the evolution of its language? Whereas early relativists emphasised vocabulary to explain differences in culture (e.g., Intuit words for snow), modern linguists now tend to study sentence structure or word placement (Lucy (1992)). In the same sense, we will take statistical relativism to mean concepts about how data are generated that are tied to the structure or form of models used to characterise those data. Clearly, this is quite natural for generative models but what about parametric models of mappings used in exploratory analyses? These sort of models are exemplified nicely by the interesting combination of singular value decomposition (SVD) and independent component analysis (ICA) presented in Bai, Shen, Huang and Truong (this issue) "to explore spatio-temporal features in fMRI data". Another example of the exploratory approach can be found in Ombao, Shao, Rykhlevskaia, Fabiani and Gratton (this issue) who "develop the concept of a location-dependent temporal spectrum for a wide class of spatio-temporal processes." This furnishes a potential data-feature that may disclose important dependencies, in relation to cognition, sleep, diagnosis and treatment outcomes.

As noted above, analyses of functional connectivity are, by definition, concerned with establishing a significant dependence or mutual information between regional activities in different parts of the brain. Bellec, Marrelec and Benali (this issue) address an interesting question about how to detect changes in functional connectivity; note that these changes "may be used to track brain reorganization while, for example, a subject learns a new skill." We will focus on two of the many developments in the field of functional connectivity, namely, resting-state functional connectivity and Granger causality.

4.1 Functional connectivity and resting-state correlations

Since the finding of Biswal, Yetkin, Haughton and Hyde (1995) that low-frequency coherence between region-specific fMRI time series recapitulates the spatial deployment of functional (motor) systems, there has been a plethora of papers analysing correlations among data acquired at rest. The advent of these analyses has been a source of contention between the proponents of resting-state functional connectivity and those trying to understand task and context-dependent changes in effective connectivity. For some, the main problem with resting-state correlation studies is that there is no hypothesis or model comparison above and beyond the question: Do the endogenous hemodynamics in one part of the brain depend on hemodynamics in another? Clearly, the spatial organisation of these dependencies can be related in an interesting way to known functional brain architectures (or profiles of metabolism) but beyond this there is no question pertaining to the mechanisms of functional integration. In particular, the linear models used to test for dependencies do not allow one to ask how connectivity changes with experimental manipulations. This contrasts with the questions about changes in effective connectivity inherent in DCM and the changes in functional connectivity considered by Bellec, Marrelec and Benali (this issue).

Resting-state functional connectivity provides a nice example of statistical relativism. Typically, data are low-pass filtered to preserve frequencies in the range 0.1 to 0.01 Hz. Simple linear models are then used to test the null hypothesis of independence between pair-wise time series from different parts of the brain. It is usually assumed that the source of this dependency, when detected, is due to slow fluctuations in underlying neuronal activity. People have speculated on the source of slow neuronal dynamics that might subtend observed low-frequency correlations. The problem with this line of thinking is that there is no generative model, based on neuronal activity, which supports this interpretation. In other words, if one generated synthetic neuronal data with broadband (e.g., scale-free) coherence that showed no frequency-specificity, and then convolved the neuronal time series with a HRF to generate synthetic fMRI data, one would see (after addition of observation noise) coherence in and only in the frequency range 0.1 to 0.01 Hz. This is because these are the only frequencies that are passed by the hemodynamic response function. In short, low-frequency correlations in hemodynamic responses do not mean that the correlations among neuronal responses are low-frequency. This example highlights the dangers of reifying a model of dependencies among data-features, in the absence of a true generative model.

4.2 Granger causality and vector autoregressive models

Granger causality represents another example of statistical relativism, in which a test for mutual information becomes reified as a causal relationship. Granger causality rests upon the rejection of the null hypothesis of statistical independence between a time series at one point of the brain and the history of activity at others. Despite its name, this falls in the class of tests for statistical dependency with no generative model. The reason that there is no underlying generative model is that the vector autoregression model used to test for statistical dependence does not model neuronal activity. In other words, inferences are made purely on the basis of observed hemodynamic responses and cannot be used to make inferences about causal interactions mediated by neuronal connections. A similar criticism could be levelled at any technique that uses hemodynamic responses to, for example, "obtain accurate temporal ordering of the various regions of the brain involved in a cognition experiment" (Lindquist, Zhang, Glover and Shepp; this issue). Perhaps the simplest example of the dangers of reifying vector autoregression (VAR) models is when the hemodynamic response function in a source area has a longer latency than in a target area. Because neuronal responses in the target area will be expressed, hemodynamically, *before* their causes in the source area, Granger causality will be inferred *from the target to the source*. In fact, the development of DCM was motivated by the failure of conventional signal processing models like VAR models to furnish proper generative models of underlying neuronal interactions.

4.3 Multivariate models for pattern classification

Another notable development over the past few years is the use of multivariate pattern classification procedures to establish statistical dependence between distributed responses in a circumscribed part of the brain and some experimental variable. These analyses have excited much attention and claim to be more sensitive than equivalent univariate analyses (e.g., Haynes and Rees (2006)). This claim is certainly true but conflates the multivariate aspect of these approaches with classification. Multivariate models are generally more sensitive than univariate models because they have less localising power, irrespective of whether one tests for a mapping between experimental factors and brain responses or *vice versa*. Classification is somewhat incidental here and just provides a surrogate for optimal tests of statistical dependency. Having said this, the cross-validation tests used in classification analyses have intuitive appeal and (like re-randomisation procedures) are robust and relatively assumption free. From the

point of view of people who have invested in generative models, multivariate pattern classification procedures are interesting but represent something of a retrograde step. This is because they represent a return to neo-phrenology: The only inference that is obtained from a multivariate classification analysis is that there is some probabilistic mapping between activity in a subset of brain voxels and some aspects of the sensorium or behaviour. Put simply, this just establishes functional specialisation, albeit of a regionally finessed and distributed sort. Although an important aspect of neuroimaging, this sort of inference is generally seen as a prelude to more mechanistic models of how specialisation emerges and the connectivity architectures that support it.

Pattern classification analyses harbour one of the more pernicious examples of statistical relativism in the rhetoric of *decoding*[2]. The supposition here is that if a statistical dependence can be established between neuronal activity and its sensory cause, one has effectively decoded the brain's encoding of that cause or attribute. This is not the case. Neuronal encoding involves the inversion of a generative model of sensory data and unless there is an explicit model of this perceptual inversion, there can be no inference about decoding. A simple example here would be a successful *decoding* of information in the visual field based on calcium imaging of the retina. This *decoding* does not mean that the retina *encodes* or perceives the causes of the sensory input, it just means that there is a statistical dependence between sensory information and the stimulus.

5. Conclusion

It might be asked, do we really need to know the neuronal mechanisms underlying fMRI responses? In the context of brain mapping *per se*, the answer is probably "no". If one was simply interested in which parts of the brain exhibit neuronal responses to particular experimental manipulations, then it is sufficient to know that fMRI signals are caused by changes in neural activity. This is because a significant change in fMRI signal must be caused by a significant change in neuronal activity. It is not necessary to know the mechanisms or nature of the causal link. However, if one is interested in the underlying dynamics and computations subtending these signals, then the mappings between different levels of description (neural codes, population

[2] A rhetoric which I personally find very appealing (e.g., Friston, Chu, Mourão-Miranda, Hulme, Rees, Penny and Ashburner (2008). Bayesian decoding of brain images . NeuroImage 39, 181-205).

dynamics, electrophysiological and hemodynamic) become critical. Over the past years, a dialectic has emerged in the imaging neuroscience community. On the one hand, there is a move towards mechanistic and biophysically plausible generative models of the signal, as exemplified in articles by Sánchez-Bornot, Martínez-Montes, Lage-Castellanos, Vega-Hernández and Valdés-Sosa (this issue) and Guha and Biswas (this issue). On the other hand, there have been developments in multivariate pattern classification and decoding approaches, which are purely descriptive or phenomenological in nature (e.g., Haynes and Rees (2006)). Both are united in a focus on the neuronal code. However, generative modelling tries to explore the space of mechanistic and algorithmic models of brain function, whereas decoding approaches are concerned with the spatial deployment of signals that may encode something in the environment. Decoding approaches can be regarded as a refinement of the brain mapping initiative that go beyond questions about where the code is located in the cortex to address the nature of its spatial and temporal distribution. In contrast, generative models try to connect brain imaging data to the underlying anatomy, electrophysiology and algorithms employed by the brain.

This review of approaches to data in imaging neuroscience is biased by a strong preference for analyses based upon generative models. There is a principled reason for this: The inversion and comparison of generative models allows one to answer questions about how the brain works. Operationally, this proceeds by comparing different generative models of observed data and assessing one model in relation to others, in terms of their marginal likelihood or evidence (the probability of the data given a model). Increasingly, over the past few years, we have seen the benefit of this approach in terms of solving some very hard problems in neuroimaging. These range from the number of equivalent current dipoles to use in distributed reconstructions of EEG data (c.f., Vega-Hernández, Martínez-Montes, Sánchez-Bornot, Lage-Castellanos and Valdés-Sosa; this issue), through to the selection among many alternative DCMs of distributed responses using fMRI (e.g., Fairhall and Ishai (2007)). Usually, this entails some form of Bayesian model comparison and a departure from classical statistics. The alternative is to use models of mappings and classical statistics or cross-validation procedures to establish statistical dependencies. This reduces to a comparison of two models, one with a mapping and one without. The only interesting information that is obtained from this sort of model comparison rests on where and when the variables came from. Although this is clearly an important endeavour, particularly in exploratory

analyses, generative models seem to meet the questions about mechanisms and functional architectures in a more direct fashion. On a final note, the articles solicited by the Guest Editors of this issue represent a healthy balance of impressive generative modelling and exploratory methods with, happily, a slight bias to the former.

References

Bandettini P. A., Jesmanowicz A., Wong E. C. and Hyde J. S (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30**, 161-73.

Biswal B., Yetkin F. Z., Haughton V. M. and Hyde J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* **34**, 537-41.

Daunizeau J., Grova C., Marrelec G., Mattout J., Jbabdi S., Pélégrini-Issac M., Lina J. M. and Benali H. (2007). Symmetrical event-related EEG/fMRI information fusion in a variational Bayesian framework. *NeuroImage* **36**, 69-87.

Fairhall S. L. and Ishai A. (2007). Effective connectivity within the distributed cortical network for face perception. *Cereb. Cortex* **17**, 2400-6.

Friston, K. J., Jezzard, P. and Turner, R. (1994). Analysis of functional MRI time-series. *Hum. Brain Mapp.* **1**, 153-171.

Friston K. J., Harrison L. and Penny W. (2003). Dynamic causal modelling. *NeuroImage* **19**, 1273-302.

Haynes J. D. and Rees G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* **7**, 523-34.

Lucy, J. A. (1992). *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis.* Cambridge University Press, Cambridge.

Marr D. (1976). Early processing of visual information. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **275**, 483-519.

Penny W. D., Stephan K. E., Mechelli A. and Friston K. J. (2004). Comparing dynamic causal models. *NeuroImage* **22**, 1157-72.

— Karl J. Friston FRS

Karl Friston is a neuroscientist and authority on brain imaging. He invented statistical parametric mapping; SPM is used internationally for analyzing functional imaging data. In 1994, his group developed voxel-based morphometry. VBM detects differences in neuroanatomy; it is also used clinically and as a surrogate in genetic studies. These technical contributions were motivated by schizophrenia research and theoretical studies of value-learning; in 1995 this work was formulated as the 'disconnection hypothesis' of schizophrenia. In 2003 he created dynamic causal modelling. DCM is used to infer the architecture of distributed systems like the brain. Friston currently works on models of functional integration in the human brain and the principles that underlie neuronal interactions. He received the first Young Investigators Award in Human Brain Mapping (1996) and was elected a Fellow of the Academy of Medical Sciences (1999) in recognition of contributions to the bio-medical sciences. In 2000 he was President of the International Organization of Human Brain Mapping. In 2003 he was awarded the Minerva Golden Brain Award, and is currently among the top ten cited scientists in neuroscience and behaviour. He was elected a Fellow of the Royal Society in 2006.