Experimental design and Statistical Parametric Mapping - Karl Friston

1. Intro

-Characterizing a regionally specific effect rests on estimation and inference
-Functional specialization and integration serve as motivation for most analyses of neuroimaging data
--These two have to be combined for full understanding of brain mapping results
-Statistical parametric mapping is generally used to identify functionally specialized brain responses
--Characterizes functional anatomy and disease-related changes
--voxel-based, classical inference, comments on regionally specific responses to experimental factors

2. Functional specialization and integration

-Brain has two fundamental principles of functional organization
--Functional integration
--Functional specialization

2a. Functional specialization and segregation
-Functional role of a brain component is defined by its cortical connections
-Functional segregation demands that cells with common functional properties are grouped together
-The analysis of functional neuroimaging data is divided into:
--Spatial processing
--Estimating parameters of a statistical model
--Inference on parameter estimates with appropriate statistics

3. Spatial Realignment and Normalization
(Section I: Computational Neuroanatomy)

-Analysis of neuroimaging data starts with series of spatial transformations
--reduce unwanted variance components in voxel time-series induced by movement or shape differences among series of scans
--voxel-based analyses assume data are derived "locally" (enabling reporting regionally-specific effects)
-First step is the realign data
--then transform using linear or nonlinear warps into a standard anatomical space
--finally, data are spatially smoothed

(Chapter 2: Rigid body registration)
3a. Realignment

-Changes in signal intensity over time arise from head motion, disrupting fMRI study results
-Realignment involves:
--estimating the 6 parameters of an affine 'rigid-body' transformation that minimizes the differences (LSA)
---first-order approximation of the Taylor expansion of the effect of movement on signal intensity using spatial derivatives
---allows for a simple iterative least squares solution corresponding to a Gauss-Newton search
--applying the transformation by re-sampling the data using tri-linear, sinc or spline interpolation.

3b. Adjusting for movement related effects in fMRI

-as much as 90% of variance in fMRI time-series can be effects of movement after realignment
--caused by effects that cannot be modeled using a linear affine model
--nonlinear effects include:
---subject movement between slice acquisition
---interpolation artifacts
---nonlinear distortion due to magnetic field inhomogeneities
---spin-excitation history effects
-these effects create movement-related signal "y", a nonlinear function of displacement "x" in current and previous scans
--y_n=f(x_n, x_(n-1),...)
-this estimated signal is then subtracted from original data
-adjustment can be carried out pre-processing step or embodied in model estimation during the analysis
-this considers spatial realignment, not temporal realignment
--temporal realignment: using sinc interpolation over time and only when:
---temporal dynamics of evoked responses are important
---TR (time repetitions) sufficiently small to permit interpolation
--timing effects are usually unimportant
--provided that effects of latency differences are modelled, this renders temporal realignment unnecessary usually

(Chapter 3: Spatial Normalization using basis functions)
3c. Spatial normalization

-After realignment, a mean image of the series is used to estimate some warping parameters that map it into a template that conforms to a standard anatomical space
-Estimation can use a variety of models for the mapping:
--12-parameter affine transformation
---parameters constitute a spatial transformation matrix
--low-frequency basis spatial functions
---discrete cosine set or polynomials
---parameters are coefficients of basis functions
--vector field specifying the mapping for each control point (eg voxel)
---parameters are vast and vector field is bigger than image
-Estimation of parameters in any case can be done through Bayesian framework, finding deformation parameters that have maximum posterior probability p(\theta|y) given data "y"
--p(\theta|y)p(y)=p(y|\theta)p(\theta)
--ie, finding deformation (most likely) given the data
--deformation can be found by maximizing probability of getting the data, assuming current estimate of deformation is true, times probability that estimate is true
--deformation is updated iteratively using Gauss-Newton scheme to maximize maximum posterior probability p(\theta|y)
---involves jointly minimizing the likelihood and prior potentials
----likelihood potential is sum of squared differences between template and deformed image
-----reflects probability of actually getting that image if the transformation was correct

----prior potential is used to incorporate prior info about likelihood of a given warp
-----can be determined empirically or motivated by constraints on the mappings
-----play a more essential role as the number of parameters increases and are central to high dimensional warping schemes
-Affine or spatial basis function warps and iterative least squares are used to minimize posterior potential

3d. Co-registration of functional and anatomical data

-Can be useful
-Distortion is not an issue if functional data is spatially normalized

3e. Spatial smoothing

-Motivations for smoothing data:
--by the matched filter theorem, the optimum smoothing kernel corresponds to the size of the anticipated effect
--by the central limit theorem, smoothing data will render errors more normally distributed and ensure validity of inferences based on parametric tests
--when inferring about regional effects using Gaussian random field theory, the assumption is that error terms are a reasonable lattice representation of an underlying and smooth Gaussian field
--in context of inter-subject averaging, often necessary to smooth more to project data onto a spatial scale where homologies in functional anatomy are expressed among subjects

3f. Summary

-Products of spatial normalization are bifold:
--a spatially normalized image and a deformation field
---deformation field contains important info about anatomy
----key part of computational neuroanatomy
-----tensor fields can be analyzed directly (deformation-based morphometry)
-----tensor fields can create maps of specific anatomical attributes (compressions, shears)
------maps can be analyzed by voxel (tensor-based morphometry)
-----normalized structural images can undergo satirical analysis (voxel-based morphometry)
------voxel-based morphometry is most common voxel-based neuroanatomical procedure

(Sections II and III: Modeling and Inference)
4. Statistical Parametric Mapping

-Statistical Parametric Mapping: the construction of spatially extended statistical processes to test hypotheses about regionally specific effects
-SPMs (maps) are image processes with voxel values that are distributed according to a known PDF, usually Student T or F
--T-maps or F-maps
-One analyzes each voxel and the resulting parameters are assembled into an image (the SPM)
-SPMs are interpreted as spatially extended processes by referring to the probabilistic behavior of Gaussian fields

--Gaussian random fields (GRF) model probabilistic characteristics of a SPM and any non-stationary spatial covariance structure
--'Unlikely' excursions of the SPM are interpreted as regionally specific effects (sensorimotor or cognitive process)
-SPM uses the general linear model (GLM) and GRF to infer data through SPMs
--GLM estimates parameters that could explain spatially continuous data
--GRF is used to resolve multiple comparison problem that ensues when making inferences over a volume of the brain
-Reason behind SPM:
--acknowledge Significance Probability Mapping, the use of interpolated pseudo-maps of p values used to summarize the analysis of multi-channel ERP (event-related potential) studies
--parametric statistics that comprise the maps
-Subtle motivations despite simplicity of method:
--mass-univariate analyses rather than multivariate analyses
---multivariate does not support inferences about regionally specific effects
---multivariate requires more observations than the dimension (number of voxels)
---in dimension reduction, multivariate approach is less sensitive to focal effects
---multivariate uses too many parameters (increasing variability of estimate of a parameter), thus inefficient
--the minimal parameterization lends SPM added sensitivity
---GRF theory implicitly imposes constraints on non-sphericity implied by the continuous and extended nature of data
-Bayesian alternative to classical inference with SPMs:
--uses Posterior Probability Maps (PPMs), less common than SPMs

4a. The General Linear Model (Chapter 7)

-Statistical analysis of imaging data corresponds to:
--modeling the data to partition observed neurophysiological responses into components of interest, confounds and errors
--making inferences about the interesting effects in relation the error variance
-the T statistic provides a more versatile and generic way of assessing the significance of regional effects and is preferred over correlation coefficient
-GLM is aka 'analysis of covariance' or 'multiple regression analysis'
--the matrix X that contains the explanatory variables is called the "design matrix"
---the column of design matrix corresponds to an effect built into the experiment (explanatory variables, covariates or regressors)
--the relative contribution of each column is assessed using standard least squares and inferences using T or F stats
-Design matrix:
--can contain both covariates and indicator variables
--each column has an associated unknown parameter (only some are of interest)
--the remaining parameters pertain to confounding effects and are not interesting
--inference about parameter estimates are made using estimated variance
---this allows testing null hypothesis (that all estimates are zero) using F stat to give SPM{F} or that a particular linear combination is zero using SPM {T}

---the T stat is obtained by dividing a contrast/compound of the ensuing parameter estimates by its standard error
----standard error of compound is estimated using variance of the residuals about the least-squares fit
-In most analysis, the design matrix contains indicator variables or parametric variables encoding the experimental manipulations
-An important instance of GLM is the linear time invariant (LTI) model
--it explicitly treats the data-sequence as an ordered time-series and enables a signal processing perspective that is useful

[1. LTI systems and temporal basis functions]

4b. Statistical inference and Random Field theory

-Classical inferences using SPMs can be of two sorts
--Anatomically constrained hypothesis
---uncorrected p value associated with the height or extent of that region in the SPM can be used to test the hypothesis
--Anatomically open hypothesis
-The theory of random fields provides a way of adjusting the p value that takes into account the fact that neighboring voxels are not independent by continuity
--For smooth data, the GRF correction is more sensitive than a Bonferroni correction
--GRF theory deals with multiple comparisons problems in the context of continuous, spatially extended statistical field
-Difference between GF and Bonferroni corrections:
--Bonferroni correction controls expected number of false positive voxels
--GRF correction controls expected number of false positive regions
--the corrected threshold under GRF is much more sensitive consequently
-Two assumptions underlying use of GRF correction:
--the error fields are a reasonable lattice approximation to an underlying random field with multivariate Gaussian distribution
--the error fields are continuous, differentiable, invertible
--assumptions are violated only if data are not smoothed (violating reasonable lattice assumption) or model is mis-specified (errors are not normally distributed)

[1. Anatomically closed hypotheses]

-Inferences about regional effects in SPMs can be predicted, but activations may want to be considered near the location
-Two approaches:
--pre-specify a small search volume and make GRF correction
--use uncorrected p value based on spatial extent of nearest cluster
-Both procedures are based on distributional approximations from GRF theory

[2. Anatomically open hypotheses and levels of inference]

-set-level inferences are generally more powerful than cluster-level inferences (more powerful than voxel-level inferences)

-price for increased sensitivity is reduced localizing power
-voxel-level tests permit individual voxels to be identified as significant
-cluster-level only allow cluster significance
-set of clusters only allow set significance
-Typically, voxel-level inferences are used and a spatial extent threshold of zero
--reflects fact that characterizations of functional anatomy are generally more useful when specified with a high degree of anatomical precision

5. Experimental Design

-Different sorts of designs in neuroimaging studies
-Experimental designs can be single-factor or multifactorial designs
--levels of each factor can be categorical or parametric

5a. Categorical designs, cognitive subtraction and conjunctions

-cognitive subtraction: two tasks are separate cognitive or sensorimotor components, thus regionally specific differences in hemodynamic responses identify functionally specialized areas
-cognitive conjunction: extension of subtraction technique, combines a series of subtractions
--conjunction tests several hypotheses, rather than just one, to see if activations, in pairs, are jointly significant
--allows demonstration of context-invariant nature of regional responses
--important in multi-subject fMRI studies

5b. Parametric designs

-parametric design: regional physiology will vary systematically with the degree of cognitive or sensorimotor processing
-neurometric functions may be linear or nonlinear
-using polynomial regression (GLM) identify nonlinear relationships between stimulus parameters (using SPM{F})
-clinical neuroscience studies use parametric designs by looking for neuronal correlation of clinical ratings over subjects

5c. Multifactorial designs

-factorial designs enable inferences about interactions
-interactions are associated with factorial designs
--the effect of one factor on another is assessed by interaction term
-interaction effects can be interpreted as:
--the integration of multiple cognitive processes
--the modulation of one perceptual process by another
-in clinical studies, interactions are central
-can also embody parametric factors
--can be expressed as a difference in regression slope of regional activity on the parameter, under both levels of the other categorical factor

6. Designing fMRI Studies
(Chapter 11: Analysis of fMRI time series)

-fMRI time-series as a linear admixture of signal and noise
--signal corresponds to neuronally mediated hemodynamic changes modeled as a convolution of some underlying neuronal process, responding to changes in experimental factors, by a hemodynamic response function (HRF)
--noise has neuronal and nonneuronal sources
---neuronal noise is neurogenic signals not modeled by explanatory variables with the same frequency structure as signal
---nonneuronal components are low frequency or wide-band
--superposition of all components induces temporal correlations among error terms that effect sensitivity to experimental effects
---sensitivity depends on:
----relative amounts of signal and noise
----efficiency of experimental design (reliability of parameter estimates, defined as inverse of variance of contrast of parameters)
-two important considerations from this perspective:
--optimal experimental design
--optimum convolution of the time-series to obtain most efficient parameter estimates

6a. The hemodynamic response function and optimum design

-LTI model of neuronally mediated signals in fMRI suggests that only experimentally induced signals that survive convolution with HRF can be estimated
-by convolution theorem the frequency structure of experimental variance should match the transfer function of HRF

6b. Serial correlations and filtering

-conventional signal processing approaches dictate that whitening the data engenders the most efficient parameter estimation
--filtering with a convolution matrix that is inverse of intrinsic convolution matrix
--the 'whitening' strategy renders the least square estimator equivalent to ML or Gauss-Markov estimator
---since the form of intrinsic correlations are unknown, must be estimated

6c. Spatially coherent confounds and global normalization

-implicit in use of high-pass filtering is removal of low-frequency components that are regarded as confounds
--also, signal components that are artifactual or have no regional specificity, called global confounds
-thus, global normalization is needed
--global estimator enters into statistical model as a confound
-in fMRI, instrumentation effects the scale data motivate global normalization before the data enter into the statistical model
-it is important to differentiate between global confounds and their estimators

6d. Nonlinear system identification approaches

-The above only considers LTI models and first order HRFs
-this signal processing perspective is by nonlinear system identification
-characterizing evoked hemodynamic responses in fMRI based on nonlinear system ID, particularly using Volterra series
--enables estimating Volterra kernels that describe relationship between stimulus presentation and hemodynamic responses
--essentially high order extensions of linear convolution models
--kernels represent nonlinear characterization of HRF modeling responses and interaction of stimuli
--in fMRI, kernel coefficients can be estimated by:
---using second order approximation to the Volterra series for GLM
---expanding kernels for temporal basis functions

6e. Event and epoch-related designs

-in experimental design, there is a crucial distinction between event- and epoch-related designs
-fMRI allows measure of event-related responses
--choice of inter-stimulus interval or SOA (stimulus onset asynchrony) is important
-designs can be stochastic or deterministic depending on whether there is a random element to their specification
--stochastic designs specify probabilities of an event occurring
--deterministic designs the event occurring is specified by stimulus
-an efficient design for one effect may not be optimal for another, even within the same experiment

7. Inferences about subjects and populations
Precision is the inverse of variance

-critical issue is whether inference is on effect related to "within-subject variability" or "between-subject"
--difference between "fixed" and "random" effect analysis
--random effects analysis allow inference to be generalized to population

7a. Random-effects analyses

-taking contrasts of parameter estimates from a "first-level" (fixed-effect) analysis and entering them into a "second-level" (random-effect) analysis
--second-level design matrix tests null hypothesis that contrasts are zero

7b. Conjunction analyses and population inferences

-motivation for conjunction analysis within multi-subject studies:
--provide inference, in fixed-effect analysis testing null hypothesis, that is more sensitive than testing average activation
--extended to make inferences about population, when conjunction of effects is established
-conjunction analysis steps:
--design matrix for explanatory variables of each experimental condition (models each subject by condition interactions)

--contrasts are specified that test for effect of interest in each subject to give series of SPM{T}
--SPM{T} are combined at a threshold to give a SPM{T_min} (ie conjunction SPM)

8. Functional Integration (Section 4)

8a. Functional and Effective connectivity
(Chapter 18: Functional integration)

-functional integration is inferred on basis of correlations among measurements of neuronal activity
-functional connectivity is correlation among remote neurophysiological events
-effective connectivity is the influence that one neural system exerts over another
--effective connectivity is dynamic (activity- and time-dependent)
--it depends upon a model of interactions
-estimation procedures employed in functional neuroimaging can be classified:
--based on linear regression models
--based on nonlinear dynamic models
-multivariate analysis are necessary to model interactions among brain regions
--inferential or data-led (exploratory)
---based on functional connectivity or covariance patterns (exploratory)
---models of effective connectivity (inferential)

8b. Eigenimage analysis and related approaches
(Chapter 19: Functional connectivity)

-most analyses of covariances among brain regions are based on singular value decomposition (SVD) of between-voxel covariances
-voxel-based PCA of neuroimaging time-series characterizes distributed brain systems implicated in sensorimotor, perceptual, or cognitive processes
--distributed systems are identified with principal components (eigenimages) corresponding to spatial modes of coherent brain activity
--simple multivariate characterization of functional neuroimaging time-series
--exploratory analysis
--PCA uses SVD to identify a set of orthogonal spatial modes for greatest variance over time
-covariance among brain regions is equal to functional connectivity
--eigenimage analysis addresses functional integration (ie connectivity)
-eigenimage analysis is limited:
--provides only a linear decomposition of neurophysiological measurements
--the set of eigenimages or spatial modes obtained is uniquely determined by constraints that are biologically implausible
-ICA (indep. comp. anal) uses entropy maximization to find, iteratively, spatial modes or dynamics that are approximately independent
--stronger requirement than orthogonality in PCA and involves removing high order correlations among modes
-Cluster analysis, voxels in a multidimensional scaling space are assigned probabilities to a small number of clusters
--characterizing temporal dynamics and spatial modes

8c. Characterizing nonlinear coupling among brain areas
(Chapter 20: Effective connectivity)

-linear models of effective connectivity assume that multiple inputs to a brain region are linearly separable
--need for models to include interactions among inputs
---these interactions (or bilinear effects) can be put in structural equation modeling using "moderator" variables that represent the interaction between two regions causing activity in a third
----modulatory effects can be modeled with nonlinear input-output models, particularly Volterra formulation
-----Volterra formulation has high face validity and biological plausibility
------its assumption is that response of a region is an analytic nonlinear function of inputs over recent past
-the influence of one region on another has two components
--direct (driving) influence of input from first (lower hierarchy) region, regardless of all other activity
---mediated by first order kernels
--activity-dependent, modulatory component that represents an interaction with inputs from the remaining (higher hierarchy) regions
---mediated by second order kernels
-context-sensitive changes in effective connectivity are most important in functional integration and have two fundamental implications for experimental design and analysis:
--experimental designs for analyses of effective connectivity are multifactorial
---because one factor is needed to evoke responses and render coupling among brain areas measurable and a second factor needs to induce change in that coupling
--models of effective connectivity embrace changes in coupling
---modeled with bilinear terms/interactions

Conclusion.

-Reviewed main components of image analysis and assessing functional integration in the brain
-key principles of functional specialization and integration were considered