Temporal, Spatial, and Spatiotemporal Models

Hao, Guanshengrui

October 24, 2012

Outline

- **Methods in Chapter 5:**
  Ignore the Temporal and Spatial correlation
  Consequences:
  1. Loss of efficiency
  2. Bias
  3. Misstatement of type I error
- **Main Point:**
  Exploit temporal correlation, spatial correlation or both
  Tradeoff:
  Higher computational costs

Emphasis

- The time series nature of the voxel observations
- Ways of exploiting this information
    1. in the Time Domain
    2. in the Frequency Domain

One Basic Model

Extend linear model from Chapter 5: $Y_t = X_t\beta + Z_t$,

- Y_t: response at time t
- X_t: desigh matrix
- Difference:
  - In linear model, ϵ is assumed as normal with mean zero and variance $\sigma^2 I$
  - Here, ϵ is replaced by Z_t of mean zero and unknown covariance structure V

One Basic Model: Continued

- General procedure: prewhitening
  Write $V = KK^T$, then $D = K^{-1}$ is called **prewhitening**
- In practice,
  V is unknown and must be estimated first

Worsley and Friston (1995)

- Original work
  To estimate the parameter β in a linear model:

$$Y = X\beta + e$$

  - Y: unsmoothed time series
  - e: error vector, containing components that are independently distributed normal with mean zero and variance σ^2

  Applied to the time series, it turns to one of allowing for **serial correlation** in a regression setting, and solutions exist

Worsley and Friston (1995): Continued

- **Special procedures**
  Smooth the time series first, then the **least squares estimator** of β becomes:

$$\hat{\beta} = (X_1^T X_1)^{-1} X_1^T KY,$$

  where$X_1 = KX$.
  Then usual least square theory produces estimates of the variance and a test statistic that can be used to assess the behavior at each voxel
- **Conclusions**
  Because of smoothing, this estimator is not fully optimal, but it's unbiased, and in many situations the loss of efficiency is not great

Bullmore et al.(1996b)

- **Idea**
  Use Trigonometric Basis functions to capture the frequency information from the time series of signal intensities
- **Model**
  For signal Y_t:

$$Y_t = \gamma \sin(\omega t) + \delta \cos(\omega t) + \gamma' \sin(2\omega t) + \delta' \cos(2\omega t) + \gamma'' \sin(3\omega t) + \delta'' \cos(3\omega t) + \alpha + \beta t + \rho_t$$

  - ω: the fundamental frequency for the data
  - $\alpha + \beta t$: linear trend
  - ρ_t: error at t, estimated via **Pseudogeneralized Least Squares**

Bullmore et al.(1996b): Continued

- **Identify active voxels**
  Follow a two-stage approach:
  - Use temporal information to calculate the Fundamental Power Quotient (FPQ)at voxel i:

$$\text{FPQ}_i = \frac{\hat{\gamma}_i^2 + \hat{\delta}_i^2}{\sqrt{2(se(\hat{\gamma}_i)^4 + se(\hat{\delta}_i)^4)}}.$$

    then find the significantly active voxels using Permutation Test
  - Assume all voxels found in the first stage are false positives and let N_{vox} be the number of those voxels in each 8-connected cluster

  Only voxels that pass a threshold for both FPQ and N_{vox} are considered active

Bullmore et al.(1996b): Continued

- **Additional detail**

  Information regarding the timing of activation can be extracted from the signs of $\hat{\gamma}$ and $\hat{\delta}$:

  - Sigh of $\hat{\gamma}$ is related to the condition of the experiment to which the voxel is responding:
    - \+ task
    - \- rest
  - Sigh of $\hat{\delta}$ is related to the timing of activation:
    - \+ anticipatory
    - \- delayed

Locascio et al.(1997)

- **Idea**
  Use traditional time series methods, Autoregressive Moving Average (ARMA) models for the fMRI time course on a voxel by voxel basis
- **CARMA**
  Called by authors since the model incorporates both contrast and ARMA components

Locascio et al.(1997): Continued

- **Model**
  At t, the signal intensity Y_t is modeled as:
  $$Y_t = \alpha_0 + \sum \alpha_i C_{it} + \beta_1 time + \beta_2 time^2 + \frac{\theta(B)}{\phi(B)} \epsilon_t,$$
  - $\sum \alpha_i C_{it}$: represents contrasts of interest between the experimental and baseline conditions
  - *time*: counts the order of successive images
  - B: the backshift operator $BX_t = X_{t-1}$
  - $\theta(B)$: Moving Average Operator
    $\theta(B) = 1 - \theta_1(B) - \cdots - \theta^q(B)^q$ for a moving average component of order q
  - $\phi(B)$: Autoregressive Operator
    $\phi(B) = 1 - \phi_1(B) - \cdots - \phi^p(B)^p$ for an autoregressive component of order p
  - ϵ_t: white noise at t

Locascio et al.(1997): Continued

- **Advantages**
  - Allow for a different model to be fit at each voxel
  - Can accommodate arbitrary experimental designs
- **Drawbacks**
  Purely temporal, not clear how one could extend their procedure to have a spatial component and still keep to the spirit

Discrete Fourier Transform

$$d_w(\omega_j) = \frac{1}{n}\sum_{k=0}^{n-1} w_k \exp(-i2\pi\omega_j\delta k)$$

- ω_j: the Fourier frequencies, $\omega_j = \frac{j\delta}{n}$
- n: the length of the time course
- δ: the sampling interval
- $j = 0, 1, \ldots, [\frac{n}{2}]$

In the frequency domain, the model for the time course:

$$d_Y(\omega_j) = d_X(\omega_j)^T\beta + d_Z(\omega_j)$$

Lange and Zeger (1997)

Procedure

- Start with time domain model $Y(t,i) = X(t,\theta_i)\beta_i + Z(t,i)$,
$$X(t,\theta_i) = \sum_{0 \le s, t-s \le T-1} \lambda(s,\theta_i) x(t-s),$$
where
  - $\lambda(\cdot,\cdot)$: two-parameter gamma family
  - $Z(t,i)$: mean zero random error
- Apply Discrete Fourier Transform to this model, it becomes:
$$d_Y(\omega_j, i) = d_X(\omega_j, \theta_i)^T \beta_i + d_Z(\omega_j, i)$$
$$d_X(\omega_i, \theta_i) = d_\lambda(\omega_i, \theta_i) d_x(\omega_j)$$
- Use an iterative algorithm of Complex Least Squares, estimates are obtained for the β and θ at each spatial location separately

Lange and Zeger (1997): Continued

Drawbacks

- Two-parameter gamma model may not be flexible enough to capture the behavior of the HRF
- There may in addition be issues of parameter identifiability and convergence of the complex least squares algorithm
- The approach is only appropriate for periodic experimental designs

Marchini and Ripley (2000)

Idea

- Start with the time domain model $Y_t = X_t\beta + Z_t$
- Assume the time series has been Preprocessed to remove trends and other confounding factors
- Then the terms $d_X(\omega_j)$ vanish except at the fundamental frequency and its harmonics

Marchini and Ripley (2000): Continued

Model

$$d_Y(\omega_j) = \begin{cases} d_X(\omega_j)^T\beta + d_Z(\omega_j) & j \in \Omega \\ d_Z(\omega_j) & \text{otherwise} \end{cases}$$

Marchini and Ripley (2000): Continued

Test statistic defined for significance of the response:

$$R_j = \frac{I(\omega_j)}{g(\omega_j)},$$

- $I(\omega_j) = n|d_Y(\omega_j)|^2$: periodogram at frequency ω_j
- $g(\cdot)$: a smoothed version of the periodogram, used as an estimator of the spectral density

For periodic designs, one needs to consider R_j only at the fundamental frequency and its harmonics

Muller et al.(2001)

Idea

- Consider a multivariate approach, in hope of teasing out, in addition to regions of activation, the functional connectivities among such regions

Muller et al.(2001): Continued

Under a periodic experimental design and focus on the fundamental frequency, consider two key parameters:

- Coherence: a measure of linear association between two time series at a particular frequency, defined as

$$\rho_{jk}(\lambda) = \frac{|f_{jk}(\lambda)|}{\sqrt{f_{jj}(\lambda) f_{kk}(\lambda)}}$$

- Phase lead: a measure of the amount of temporal displacement in the BOLD response for one region relative to another, defined as the $\nu_{jk}(\lambda)$

$$f_{jk}(\lambda) = |f_{jk}(\lambda)| e^{i\nu_{jk}(\lambda)}.$$

  where $f_{jk}(\lambda)$ is the cross-spectral density function at frequency λ

Muller et al.(2001): Continued

Advantages

- Pick out similar regions of activation compared with the standard general linear model analysis
- Some understanding of the network, via the different lags in BOLD response for different regions, is obtained

Drawbacks

- It's still not possible from this approach to infer causality in the network
- The experimental design needs to be periodic or nearly so

Gonzalez Andino et al.(2000)

Idea

- Start from the assumption that time series for voxels that are related to "signal" should look different from those that are related to "noise"
- The time series should be differentiable according to their complexity, with "signal" voxels having less complex patterns made up of a few temporal components

Gonzalez Andino et al.(2000): Continued

A Basic Measure: TFR

Time frequency representation, a two-dimensional plot showing how the frequency of a series varies over time

- Time series contain organized signal will have a few "hot spots" in the TFR related to "noise"
- Those are essentially noise will have many such spots scattered at random

Thus, the number of hot spots in the TFR can be taken as a measure

Gonzalez Andino et al.(2000): Continued

Formal Measure: Renyi entropy

$$H_\alpha(C_s) = \frac{1}{1-\alpha}\log_2 \int\int \left(\frac{C_s(t,f)dtdf}{\int\int C_s(t,f)dtdf}\right)^\alpha$$

- α: order of the entropy
- $C_s(t,f)$: the coefficients of the TFR of the time series s

Based on empirical studies, $\alpha = 3$ is chosen

Gonzalez Andino et al.(2000): Continued

Advantages

- There's no need to estimate the HRF, nor to assume a reference vector
- It can be applied to event-related experiments of arbitrary complexity

Gonzalez Andino et al.(2000): Continued

Drawbacks

- Users need to choose the time frequency representation and the order α
- It's not clear how sensitive conclusions are to these choices
- The authors do not offer a formal way of distinguishing between voxels to be declared active and inactive based on values of $H_\alpha(C_s)$ when there's not a natural separation between the two groups

Purdon and Weisskoff (1998)

- **Study**
  Reported a simulation study that explores the importance for precise statistical inference of accounting for the temporal correlation in the fMRI time series.
- **Conclusion**
  the result shows that if there're indeed temporal autocorrelations, ignoring them introduces bias in the assumed significance levels

Woolrich et al. (2001)

- **Study**
  examine the effectiveness of several statistical methods for directly handling autocorrelation in the fMRI time series.
- **Conclusion**
  Among the coloring with a low-pass filter, correcting the variance and prewhitening, the authors report that prewhitening is the most efficient method; and among the parametric or nonparametric techniques or windowing/tapering methods which are used to estimate the autocorrelation, they report that a simple windowing works best.

Reasons for Lack of Purely Spatial Models

- Physical location alone is not enough to describe the spatial dependence
- Spatial analysis that ignores temporal element still requires some prior processing

Hartvig and Jensen (2000)

Idea

- Their analysis start with the intuitively pleasing idea that active voxels will tend to cluster together. Thus it makes sense to consider activation status in clusters or neighborhood of voxels

Hartvig and Jensen (2000): Continued

Procedures

- Specify a likelihood for this observed value (get from interim statistical map), called x, given the activation status A of the voxel
- prior is specified
- Bayes rule gives the posterior probability of a pattern of activation and of a particular voxel being acive

Posterior probability for entire pattern of activation:

$$P(A_C = a_C | x_C) \propto f(x_C | a_C) P(A_C = a_C)$$

Posterior probability of voxel i being active:

$$P(A = a | x_C) \propto \sum_{a^1=0,1} \cdots \sum_{a^k=0,1} P(A_C = a_C | x_C).$$

Hartvig and Jensen (2000): Continued

Main focus: Prior

- Hartvig and Jensen (2000) propose three priors:

$$P(A_C = a_C) = \begin{cases} q_0 & S = 0 \\ q_1 & S > 0 \end{cases}$$

$$P(A_C = a_C) = \begin{cases} q_0 & S = 0 \\ \alpha\gamma^{S-1} & S > 0 \end{cases}$$

$$P(A_C = a_C) = \begin{cases} q_0 & S = 0 \\ \alpha_1\gamma_1^{S-1} + \alpha_2\gamma_2^{S-k} & 1 \le S \le k \\ q_1 & S = k+1 \end{cases}$$

where S is the number of 1s in the cluster under consideration

Hartvig and Jensen (2000): Continued

Conclusion:

- The authors recommend the second one, applied to a small neighborhood, as it performs best in terms of power and of minimizing classification error and also reduces the computational burden.

Outline

Though it's the most natural way to handle functional neuroimaging data, it's faced with both computational and conceptual barriers.

- Computationally, the task of fitting a fully spatiotemporal model to a fMRI dat is formidable one
- Conceptually, the spatial correlation in particular is difficult to summarize in a form that admits a simple statistical model

Two school of thought

- Clustering of time series
- Direct modeling

A few Questions

- What should be clustered — the raw time series or some function of these?
- What clustering algorithm or family of algorithms should be used?
- How many clusters are needed and how should this be decided?

A few Questions: Continued

What should be clustered?

- Clusters the time series, looking for similarities in behavior, (Baumgartner et al., 1997;Baumgartner et al., 1998) which is the dominant approach
- Other authors (Goutte et al., 1998), claim that clustering the time series is un stable. They recommend clustering instead on the correlation function of the series with the experimental paradigm

Also, most researchers in this area recommend doing some sort of screening first, to eliminate voxels that are clearly no active.

A few Questions: Continued

Clustering Algorithm

- Most popular algorithms: K means (Balslev et al. 2002) and fuzzy clustering (Baumgartner et al. 1998; Fadili et al. 2000)
- Hierarchical methods (Stanberry et al., 2003)
- Combination of K means and hierarchical clustering (Filzmoser et al.1999)

A few Questions: Continued

Number of Clusters

- Hierarchical clustering methods such as average, single, or complete linkage do not require prior specification of the number of clusters, although a clustering threshold must be picked
- Fuzzy clustering and K means require the number of clusters be chosen ahead of time

Algorithm to decide the number

- Two-stage approach of Filzmoser et al.(1999)
- Posteriori validation of detected clusters by statistical testing (Baumgartner et al., 1998)
- Cross-validation (Balslev et al., 2002)

MST

- Minimal spanning trees (MST) is a multidimensional generalization of an ordered list, which can be served as a means of investigating the temporal evolution and connectivity among groups of spatially clustered voxels.
- The starting point of such an investigation is therefore voxel time series that have already been clustered by some other method.

Baumgartner et al.(2001)

Definitions

- **Root node:** The root node has depth of zero, and the depths of the other time courses are defined by their distances from the root
- **Run:** A run is a consecutive sequence of voxels from the same clusters

Baumgartner et al.(2001): Continued

Total number of runs and the length of the longest run give information about the separability of the clusters:

- Many short runs: Not separable
- Few long runs: Should be separated

Coactivation is inferred when no such separation results from the MST ranking of the time courses

Baumgartner et al.(2001): Continued

One example:

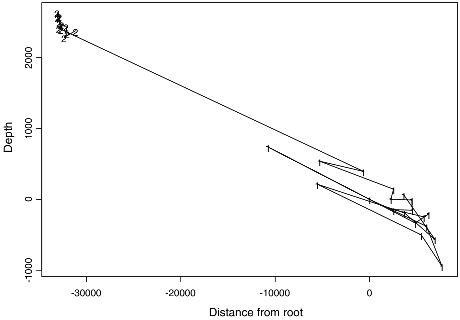

Fig. 6.1. Minimal spanning tree built from voxels belonging to two pre-identified clusters. The first cluster contains 19 voxels and is thought by the researcher to be related to the task. The second cluster contains 12 voxels and is believed to be noise. While the complete separation of the clusters apparent in the tree cannot validate these claims, it does confirm that the time courses of the voxels in the two clusters exhibit very different behaviors.

Baumgartner et al.(2001): Continued

One example (Continued):

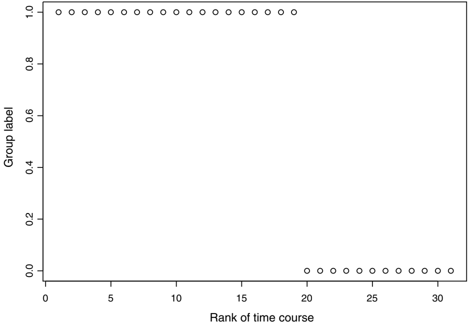

Fig. 6.2. Ordered index plot, based on the minimal spanning tree for the two clusters. The ordering is based on distances between time courses. The voxels in the two clusters are completely separated from each other.

Baumgartner et al.(2001): Continued

One example (Continued):

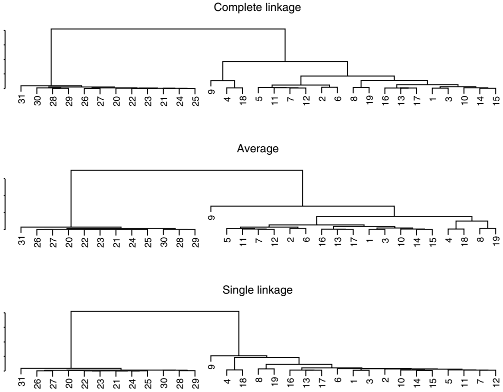

Fig. 6.3. Hierarchical clustering applied to the combined data set of $n = 31$ voxels. All three methods clearly and correctly identify the two clusters from which the voxels are drawn.

Stanberry et al. (2003)

Idea

- Use the idea of the MST, through its connection with the single linkage hierarchical analysis and "dendrogram sharpening"

Dendrogram Sharpening

- A way of reducing the data that are input to the cluster algorithm in order to produce more distinct clusters

Stanberry et al. (2003): Continued

- **Procedure of Dendrogram Sharpening**
  1. Looking at every parent node in an initial dendrogram based on all of the data, starting at the root node
  2. Any branch of the dendrogram that has a minimum number of nodes (n_{core}) is a candidate for sharpening
  3. Child nodes that are smaller that a preset size (n_{fluff}) are eliminated

Also, additional data reduction is achieved by a prescreening step prior to the sharpening

Stanberry et al. (2003): Continued

One example:

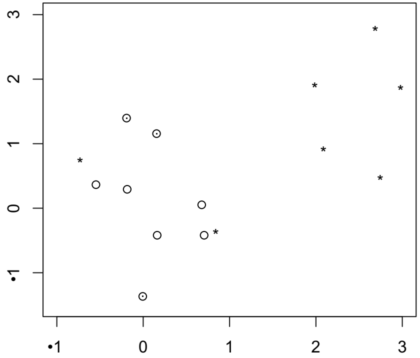

Fig. 6.4. Scatterplot of simulated data; 8 points (open circles) are taken from the standard bivariate normal and 7 (stars) from the bivariate normal with mean $(1.5, 1.5)$ and covariance matrix I. Two clusters are discernible in the data, with some amount of overlap. The three open circles with dots inside of them are points that are discarded by the sharpening algorithm.

Stanberry et al. (2003): Continued

One example (Continued):

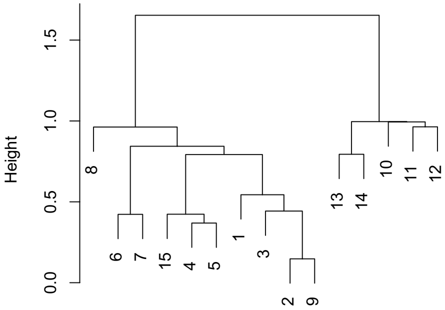

Fig. 6.5. Dendrogram for simulated data, using single linkage algorithm. Two clusters are identified; however, not all observations are classified correctly.

Stanberry et al. (2003): Continued

One example (Continued):

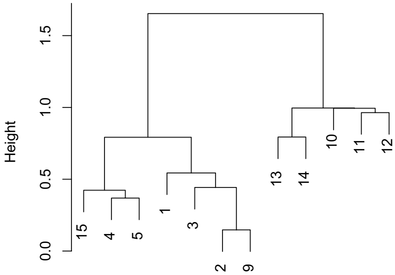

Fig. 6.6. Dendrogram of sharpened data set. The observations are numbered according to their original indices.

Direct Modeling

- From a statistical perspective it is, perhaps, the most complete and correct. But due to the computational complexity, it has only become feasible relatively recently.
- Within the rubric of "direct models" are included models based on regression and wavelets, and Bayesian models, among others.

Purdon et al.(2001)

Idea

- Seek the relationship between the input (a sensory or cognitive stimulus) and the observed output (the measured fMRI response to that stimulus)

Purdon et al.(2001): Continued

Model: comprises three components

- hemodynamic response: follows known or assumed patterns
- noise part, with two components:
  - White for the scanner noise
  - AR(1) for the low-frequency physiological noise
- drift part: linear in time, accounts for slow drift in the external field, as well as amounts of motion that were not corrected in any motion correction step

Purdon et al.(2001):Continued

At each voxel, noise parameters and signal parameters need to be estimated. The authors set an overall fitting criterion that is separable:

$$J(\theta) = \sum_{v=1}^{V} J_v(\theta_v);$$

where θ_v is the entire vector of parameters at voxel v

Purdon et al.(2001): Continued

The criterion at a given voxel v is a spatially locally weighted log-likelihood:

$$J_v(\theta_v) = \sum_{q \in N_v} K^h_{v-q} L_q(\theta_v)$$

- $L_q(\theta_v)$: Gaussian log-likelihood based on time series at voxel q
- K: Kernel function
- h: size of neighborhood on which kernel function is concentrated
- N_v: the neighborhood of voxel v

Finally, estimation proceeds iteratively, alternating between the noise and signal parameters, thus these are separately spatially regularized

Purdon et al.(2001): Continued

Advantages

- On both simulated and real data, local regularization is able to better estimate both the signal and the noise
- Estimates of noise are smoother under this analysis, while estimates of the signal aren't blurred

McIntosh et al. (2004)

Idea

- Use partial least squares specifically for event-related fMRI experiments (suitable for block designs as well)

Partial Least Squares

- A multivariate extension of multiple linear regression, where "partial" refers to computing the best least squares fit but only to part (here is related to the experimental design or to subject behavior)of a covariance (correlation matrix)

McIntosh et al.(2004): Continued

Comparison with other multivariate extensions
Consider a multiple linear regression in general form:

$$Y = X\beta + \epsilon$$

Whereas most of other methods (discriminant analysis, principal components analysis (PCA), canonical correlation analysis (CCA)) extract factors from the $Y^T Y$ or $X^T X$ matrices only, partial least squares extract factors from $Y^T Y X^T X$

McIntosh et al.(2004): Continued

Procedure of ST-PLS

- Rearrange the data array into a matrix to reflect the multivariate nature of the PLS approach:
  - Both spatial and temporal information in the columns
  - Information about the experimental design in rows
- Singular value decomposition on the rearranged matrix

McIntosh et al.(2004): Continued

Two version of ST-PLS

1. Data matrix is mean centered and factors are extracted by applying a singular value decomposition to this new matrix
2. Original data matrix is transformed by a set of orthonormal contrasts representing effects of interest. The covariance matrix of these contrasts is then calculated and the singular value decomposition is applied to this matrix instead

McIntosh et al.(2004): Continued

Result of singular value decomposition

- A set of factors relate brain activity and experimental design due to the layout of the rearranged data matrix

McIntosh et al.(2004): Continued

Validation

Two type of tasks:

- Visual processing
- auditory processing

Two types of ST-PLS:

- Task analysis to detect spatiotemporal patterns in the stimulus response
- Behavior analysis to examine the spatiotemporal structure of brain behavior and reaction time on the tasks

McIntosh et al.(2004): Continued

Validation (Continued)

Two significant factors are found in the first analysis

1. The first is attributed to the main effect of task versus rest
2. The second significant factor yields the interaction between type of stimulus (auditory or visual) and condition (task versus baseline)

One significant factor is discovered in the second analysis

- Interpreted as the overall correlation of reaction time with brain activation in both tasks

Bibliography

Balslev, D., Nielsen, F. A., Frutiger, S. A., Sidtis, J. J., Christiansen, T. B., Svarer, C., Strother, S. C., Rottenberg, D. A., Hansen, L. K., Paulson, O. B., and Law, I. (2002). "Cluster analysis of activity-time series in motor learning." *Human Brain Mapping*, **15**, 135–145.

Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., and Moser, E. (1997). "Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: Reproducibility." *Journal of Magnetic Resonance Imaging*, **7**, 1094–1101.

Baumgartner, R., Somorjai, R., Summers, R., and Richter, W. (2001). "Ranking fMRI time courses by minimum spanning trees: Assessing coactivation in fMRI." *NeuroImage*, **13**, 734–742.

Baumgartner, R., Windischberger, C., and Moser, E. (1998). "Quantification in functional magnetic resonance imaging: Fuzzy clustering vs. correlation analysis." *Magnetic Resonance Imaging*, **16**, 115–125.

Bibliography (Continued)

Bullmore, E. T., Rabe-Hesketh, S., Morris, R. G., Williams, S. C. R., Gregory, L., Gray, J. A., and Brammer, M. J. (1996b). "Functional magnetic resonance image analysis of a large-scale neurocognitive network." *NeuroImage*, **4**, 16–33.

Fadili, M. J., Ruan, S., Bloyet, D., and Mazoyer, B. (2000). "A multistep unsupervised fuzzy clustering analysis of fMRI time series." *Human Brain Mapping*, **10**, 160–178.

Filzmoser, P., Baumgartner, R., and Moser, E. (1999). "A hierarchical clustering method for analyzing functional MR images." *Magnetic Resonance Imaging*, **17**, 817–826.

Bibliography (Continued)

Gonzalez Andino, S. L., Grave de Peralta Menendez, R., Thut, G., Spinelli, L., Blanke, O., Michel, C. M., Seeck, M., and Landis, T. (2000). "Measuring the complexity of time series: An application to neurophysiological signals." *Human Brain Mapping*, **11**, 46–57.

Goutte, C., Toft, P., Rostrup, E., Nielsen, F. A., and Hansen, L. K. (1999). "On clustering fMRI time series." *NeuroImage*, **9**, 298–310.

Hartvig, N. V. and Jensen, J. L. (2000). "Spatial mixture modeling of fMRI data." *Human Brain Mapping*, **11**, 233–248.

Lange, N. and Zeger, S. L. (1997). "Non-linear Fourier time series analysis for human brain mapping by functional magnetic resonance imaging (with discussion." *Applied Statistics*, **46**, 1–29.

Locascio, J. J., Jennings, P. J., Moore, C. I., and Corkin, S. (1997). "Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging." *Human Brain Mapping*, **5**, 168–193.

Bibliography (Continued)

Marchini, J. L. and Ripley, B. D. (2000). “A new statistical approach to detecting significant activation in functional MRI.” *NeuroImage*, **12**, 366–380.

McIntosh, A. R., Chau, W. K., and Protzner, A. B. (2004). “Spatiotemporal analysis of event-related fMRI data using partial least squares.” *NeuroImage*, **23**, 764–775.

Müller, K., Lohmann, G., Bosch, V., and von Cramon, D. Y. (2001). “On multivariate spectral analysis of fMRI time series.” *NeuroImage*, **14**, 347–356.

Purdon, P. L., Solo, V., Weisskoff, R. M., and Brown, E. N. (2001). “Locally regularized spatiotemporal modeling and model comparison for functional MRI.” *NeuroImage*, **14**, 912–923.

Purdon, P. L. and Weisskoff, R. M. (1998). “Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI.” *Human Brain Mapping*, **6**, 239–249.

Bibliography (Continued)

Stanberry, L., Nandy, R., and Cordes, D. (2003). "Cluster analysis of fMRI data using dendrogram sharpening." *Human Brain Mapping*, **20**, 201–219.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001). "Temporal autocorrelation in univariate linear modeling of FMRI data." *NeuroImage*, **14**.

Worsley, K. J. and Friston, K. J. (1995). "Analysis of fMRI time-series revisited– Again." *NeuroImage*, **5**, 173–181.

— (2000). "A test for a conjunction." *Statistics and Probability Letters*, **47**, 135–140.

THANK YOU