

chapter eight

Application of change point theory to modeling state-related activity in fMRI

Martin A. Lindquist
Columbia University

Tor D. Wager
Columbia University

Contents

Introduction	150
Methods	153
Exponentially weighted moving averages	153
EWMA statistic variance	157
Correction for search across time	157
Estimating change points	158
Dealing with multiple change points	159
HEWMA: A hierarchical extension of EWMA	160
Simulations	163
Simulation 1	163
Simulation 2	164
Simulation 3	165
Experimental fMRI data collection and analysis	165
Participants	165
Task design	165
Image acquisition	166
Image analysis	166
Results	168
Simulation results	168
Simulation 1	168

Simulation 2	168
Simulation 3	169
fMRI results	170
Discussion	174
Conclusion	177
Acknowledgment	178
References	178

Introduction

Neuroimaging, particularly functional magnetic resonance imaging (fMRI), is a useful and increasingly popular tool for studying the brain activity dynamics underlying human psychology. fMRI data consist of brain images — approximately 100,000 brain *voxels* (cubic volumes that span the three-dimensional space of the brain) — measured repeatedly every 2 sec or so, with images nested within task conditions and conditions nested within participants. Analysis of these complex data sets is the focus of intensive research and development in the scientific community.

Most previous approaches toward analysis assumed that the nature, timing, and duration of the psychological processes are known (Wager, Hernandez, Jonides, & Lindquist, 2007; Worsley & Friston, 1995). The analysis thus tests whether activity in a brain region (i.e., a voxel or set of voxels) is systematically related to the known input function. In these model-based approaches, the general linear model (GLM) is used to test for differences in activity among psychological conditions or groups of participants. For example, a task might require participants to inhibit an automatic tendency to make a motor response in favor of a less automatic one. The investigators might look for transient increases in brain activity when inhibition-demanding stimuli are presented and test whether activity is greater on trials that require inhibition compared with those that do not.

However, in many areas of psychological inquiry, the precise timing and duration of psychological events is hard to specify *a priori*. For example, a participant watching a funny film might experience amusement that builds gradually over time. The time course of the psychological process is likely to vary idiosyncratically from participant to participant, and without moment-by-moment reports, the progression of amusement over time will be difficult to specify. The need for obtaining a valid and accurate measure of psychological activity to use as a predictor for brain activity is a limitation as valid psychological indices are not always available. Self-reported emotion, for example, may not give an accurate picture of emotional experience in many situations. In addition, reporting emotions might change the underlying experience in important ways

(Taylor, Phan, Decker, & Liberzon, 2003). Thus, our ability to localize emotional activity in the brain is only as good as our ability to accurately tell what someone is feeling through other means. Brain activity may provide a *better* measure of some aspects of emotional processing, but this information is difficult to uncover in the linear modeling system because, to the degree that brain activity departs from a priori predictions based on behavioral measures, it will tend to go undetected.

Much of the literature on emotion has attempted to partially circumvent the problem raised by the uncertain time course corresponding to emotion by presenting brief emotionally evocative events of different types and looking for activity increases that follow a specified canonical model (Hariri, Mattay, Tessitore, Fera, & Weinberger, 2003; Lang et al., 1998; Ochsner et al., 2004; Wager, Phan, Liberzon, & Taylor, 2003). The cost of this approach is that these studies cannot investigate strong, naturalistic emotions. Many other processes are similarly difficult to assess psychologically, including anticipatory processes, shifts in strategy or performance, and changes in states of arousal.

A related problem is that even in tasks with a psychological profile that is relatively easy to specify, the specification relies on logical task analysis and may not map precisely onto brain function. For example, an attention-switching task might require participants to switch attention between two objects on some trials but not others (Wager, Jonides, Smith, & Nichols, 2005; Yantis et al., 2002). The knowledge of which trials are “switch trials” is used to build the GLM model, but the model cannot capture spontaneous shifts in attention that are not proscribed by the task instructions. Thus, there is a need to develop analyses appropriate for modeling activity with a psychological time course that is uncertain.

An alternative analysis strategy that may be appropriate in such situations is the data-driven approach. Rather than taking psychological activity as given and testing for brain changes that fit the psychological model, data-driven approaches attempt to characterize reliable patterns in the data and relate those patterns to psychological activity post hoc. One particularly popular data-driven approach in the fMRI community is independent components analysis (ICA), a variant on a family of analyses that also includes principal components and factor analyses (Beckmann & Smith, 2004; Calhoun, Adali, & Pekar, 2004; McKeown & Sejnowski, 1998). Recent extensions of these methods can identify brain activity patterns (components) that are systematic across participants (Beckmann & Smith, 2005; Calhoun et al., 2004) and can identify state-related changes in activity that can subsequently be related to psychological processes. However, these methods do not provide statistics for inferences about whether a component varies over time and when changes occur in the time series. In addition, because they do not contain any model information, they

capture regularities whatever the source; thus, they are highly susceptible to noise, and components can be dominated by artifacts.

The development of hybrid data and model-driven approaches has the potential to mitigate some of the worst problems of each approach (Calhoun, Adali, Stevens, Kiehl, & Pekar, 2005; McIntosh, Chau, & Protzner, 2004). In this chapter, we discuss a hybrid approach for identifying changes in fMRI time series in individual and group data that allows for valid population inference. The approach uses ideas from statistical control theory and change point (CP) theory to model slowly varying processes with uncertain onset times and durations of underlying psychological activity. The main benefits of the methods presented here are that they are semi-model-free methods of detecting activation and are therefore insensitive to the phase lag of the hemodynamic response. In this sense, they share some of the attractive features of data-driven analysis methods. However, on the other hand, they retain the inferential nature of the more rigid modeling approach.

The change point analysis that we develop is a multisubject extension of the exponentially weighted moving average (EWMA) change point analysis (also called statistical quality control charts; Neubauer, 1997; Roberts, 1959; Shehab & Schlegel, 2000). Activity during a baseline period is used to estimate noise characteristics in the fMRI signal response. This activity is used to make inferences on whether, when, and for how long subsequent activity deviates from the baseline level. We extend existing EWMA models for individual subjects (a single time series) to include autoregressive process of order p (AR(p)) and autoregressive moving average process (ARMA(1,1)) noise processes, then develop a group analysis using a hierarchical model, which we term HEWMA (hierarchical EWMA). Group-level changes are of primary interest in the neurosciences because group analyses can be used to make inferences about how a population of participants perform a task (often referred to as second-level or “random effects” analyses in the fMRI literature; Wager et al., 2007). Finally, we apply the method to the detection of differences between groups (e.g., patients vs. controls) or conditions within a group.

Once a systematic deviation from baseline has been detected in the group, we estimate the time of change and recovery time (if any). Variations across the brain in the onset and number of CPs and in the duration of a shift away from baseline activity can be used to constrain inferences about the functions of brain regions underlying changes in psychological state. CPs are of potential interest for a number of reasons. First, they may provide a basis for discriminating anticipatory activity from responses to a challenge (e.g., activity that begins in anticipation of pain from that elicited by painful stimuli; Koyama, McHaffie, Laurienti, & Coghill, 2005;

Wager, 2005). Second, CPs may be used to identify brain regions that become active at different times during a challenge (e.g., the early, mid-, or late phases of a tonic painful stimulus). Third, CPs may provide meaningful characterizations of differences among individuals. In clinical studies, the onset time of brain responses to anxiety may provide clinically relevant markers of anxiety disorders. In studies of emotion, the speed of recovery from adverse events is thought to be an important predictor of emotional resilience (Fredrickson, Tugade, Waugh, & Larkin, 2003; Tugade & Fredrickson, 2004), and CPs could provide direct brain measures of recovery time. In cognitive psychology, brain CPs in problem solving and insight tasks may provide a direct neural correlate of traditional time-to-solution measures in cognitive studies (Cheng & Holyoak, 1985; Christoff et al., 2001).

The HEWMA method could be used to analyze fMRI data voxelwise throughout the brain, data from regions of interest (ROIs), or temporal components extracted using ICA or similar methods. Here, we provide power and false-positive rate (FPR) analyses based on simulations, and we apply HEWMA to voxelwise analysis of an anxiety-producing speech preparation task. We demonstrate how the method detects deviations from a pretask instruction baseline and can be used to characterize differences between groups in both evoked fMRI activity and CPs.

Methods

There is a large variety of CP detection problems that present themselves in the analysis of time series and dynamical systems. In this chapter, we apply ideas from statistical control theory and CP theory to model slowly varying brain processes (i.e., periods of enhanced neural activity detectable with fMRI) with uncertain onset times and durations. We first develop the method for detecting CPs in a single time series using EWMA and then develop a hierarchical extension (HEWMA) appropriate for multisubject fMRI studies.

Exponentially weighted moving averages

Given a process that produces a sequence of observations $X = (x_1, x_2, \dots, x_n)$. (i.e., an fMRI time series), we first consider a two-state model in which the data are modeled as the combination of two normal distributions, one with mean θ_0 and (co)variance Σ , and the second with mean θ_1 and covariance Σ . During a baseline acquisition period, the process generates a distribution of data with mean θ , and while in this state the process is considered to be *in control*. The observations follow this distribution up

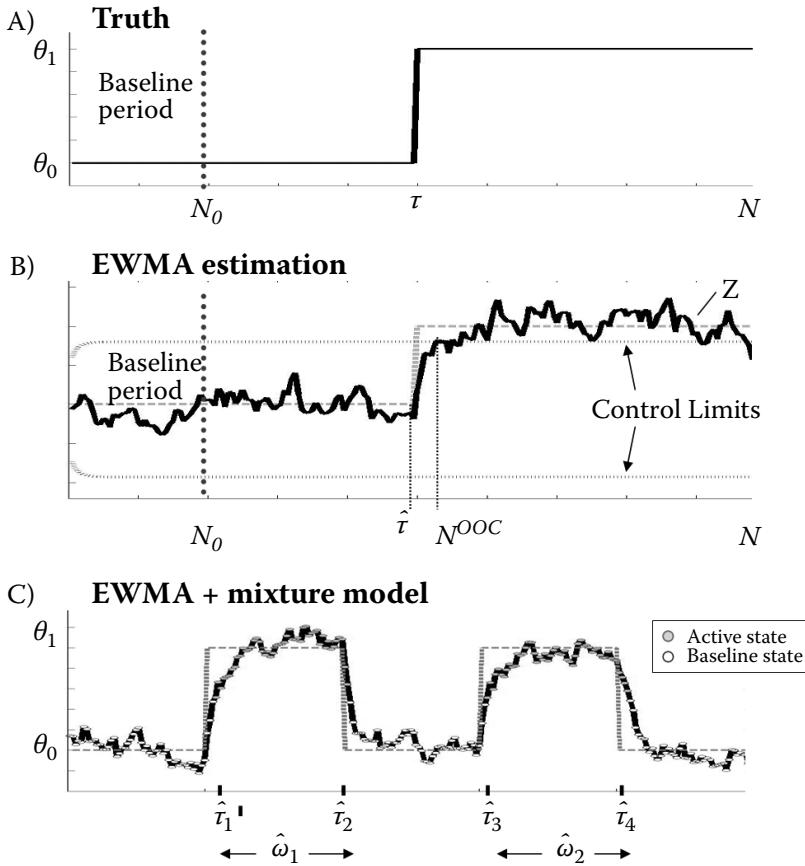


Figure 8.1 A schematic overview of: (A) the model of true activation, (B) the EWMA statistic and its control bounds, and (C) the Gaussian mixture model.

to some unknown time τ , the CP, when the process changes (i.e., a new psychological state results in increased or decreased neural activity), resulting in the generation of fMRI observations from the second distribution with mean θ_1 (see Figure 8.1A). While in this second state, the process is deemed to be *out of control* or in the out-of-control (OOC) state. The statistical model for this framework can be written as follows:

$$x_t = s_t + \varepsilon_t \quad \text{for } t = 1, \dots, n \quad (8.1)$$

where s_t denotes the signal, and ε_t is the error at time t . s_t is specified by

$$s_t = \begin{cases} \theta_0 & \text{for } t = 1, \dots, \tau \\ \theta_1 & \text{for } t = \tau + 1, \dots, n \end{cases} \quad (8.2)$$

and ε_t is specified by a mean-zero normal distribution with covariance matrix Σ , that is,

$$\varepsilon_t \sim N(0, \Sigma) \quad (8.3)$$

The diagonals of the $n \times n$ matrix Σ are the error variance estimates for each observation, and the off diagonals are the covariance among observations induced by autocorrelation in the fMRI time series. At a later stage, we relax the constraint of a single CP.

The EWMA control chart is based on the EWMA statistic z_t . The statistic is a temporally smoothed version of the data, which provides some regularization of the data and increases the power of the ensuing statistical test. It is defined as follows:

$$z_t = \lambda x_t + (1 - \lambda)z_{t-1} \quad \text{for } t = 1, \dots, n \quad (8.4)$$

or, equivalently,

$$z_t = \lambda \sum_{j=0}^{t-1} (1 - \lambda)^j x_{t-j} + (1 - \lambda)^t \theta_0 \quad \text{for } t = 1, \dots, n \quad (8.5)$$

where $0 < \lambda < 1$ is a constant smoothing parameter chosen by the analyst, and the starting value z_0 is set equal to an estimate of the process target (e.g., the baseline mean θ_0). Thus, each value of z_t is a weighted average of the current observation x_t and the previous value of the EWMA statistic. Notably, since the EWMA statistic is a weighted average of the current and all past observations, it is relatively insensitive to violations to the normality assumption.

Smoothing the data can increase power to detect deviations from the null model by regularizing the data; however, the optimal choice of smoothing parameter depends on the nature of the deviations. A general rule of thumb is to choose λ to be small (more smoothing) if one is interested in detecting small but sustained shifts in the process and larger (less smoothing) if the shifts are expected to be large but brief. Optimal values of λ are given by Lucas and Saccucci (1990).

In general, we are interested in making inferences on two features of the model. First, we seek to develop statistical tests to determine *whether* a change in distribution has indeed taken place (i.e., whether we can reject the null hypothesis $\theta_0 = \theta_1$). If a change is detected, we would also like to estimate *when* exactly the change took place, that is, estimate the unknown parameter τ .

For detecting activation, the null hypothesis is that there is a single, baseline state. The alternative is that there are two or more states with different mean activity levels, although here we consider only the simplest two-state (baseline and activation) alternative. Thus,

$$H_0 : \mu = \theta_0 \quad \text{for } t = 1, \dots, n$$

$$H_a : \mu = \theta_0 \quad \text{for } t = 1, \dots, \tau \quad \text{and} \quad \mu = \theta_1 \quad \text{and} \quad t = \tau + 1, \dots, n$$

Our aim is to assess the probability of observing the data under the null hypothesis $X \sim N(\theta_0, \Sigma)$. Inferences about EWMA statistic deviations from baseline can be made using a classical hypothesis testing framework, with a few adjustments described below. For each time point following the baseline period, we compute a test statistic T_t :

$$T_t = \frac{|z_t - \theta_0|}{\sqrt{\text{Var}(z_t)}} \quad (8.6)$$

where $\text{Var}(z_t)$ is the variance of the EWMA statistic at time t , the calculation of which is described below. Under the null hypothesis, this statistic follows a t distribution with df (degrees of freedom), which are based on the number of observations in the baseline period and corrected for autocorrelation using the Satterthwaite approximation. Thus, classical p values for the magnitude of each deviation from baseline may be obtained. If the p value is smaller than the desired Type I error level (e.g., .05), then the single-state null hypothesis is rejected. Confidence bands, or control limits, for EWMA statistic deviations can be calculated as follows:

$$|z_t - \theta_0| > c\sqrt{\text{Var}(z_t)} \quad (8.7)$$

where c is a critical value from the t distribution corresponding to the desired FPR control, and $\text{Var}(z_t)$ is the variance of the EWMA statistic at time t (see Figure 8.1B). Two problems must be solved to implement this test: First, $\text{Var}(z_t)$ must be formulated, and second, a value of c must be chosen that provides appropriate correction for search across n observations. We describe solutions to these problems below.

EWMA statistic variance

Using classical inference to test for deviations in the vector of EWMA statistics Z requires that we know the variance of Z . In most statistical control problems, the errors in the original data X are considered to be independent and identically distributed (*iid*), that is, $\Sigma = I\sigma^2$, where I denotes the identity matrix and σ^2 the error variance. In this case, it can be shown (Montgomery, 2000) that

$$\text{Var}(z_t) = \sigma^2 \frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2t}) \quad (8.8)$$

and as $t \rightarrow \infty$, the asymptotic variance approaches

$$\text{Var}(z_t) = \sigma^2 \frac{\lambda}{2 - \lambda} \quad (8.9)$$

However, noise in fMRI data is autocorrelated (Aguirre, Zarahn, & D'Esposito, 1997; Wager et al., 2007), so the application to fMRI analysis requires extension of the noise model to allow for nonwhite noise (WN). When the process error follows an AR(1) model, the form of $\text{Var}(z_t)$ has previously been derived (Schmid, 1997). In appendix A, we derive similar results for both the general AR(p) and ARMA(1,1) noise models. Knowledge of the exact form of the variance of the EWMA statistic allows for better and more accurate tests and simplifies the hierarchical weighting later in the group analysis. The ARMA(1,1) model is, with proper choice of constants, equivalent to the AR(1) plus WN model often used in fMRI analysis (e.g., in SPM2 software; Friston et al., 2002).

Correction for search across time

The procedure outlined above provides p values for the null hypothesis test that the process is in control at each time point. However, because the test is repeated over n observations, correction for searching over time is warranted unless there is a strong a priori prediction of which observations should be out of control. A simple correction may be obtained using Bonferroni correction for the number of observations, but this correction will be conservative (and thus not very powerful) because fMRI data are typically positively correlated across time.

A more accurate result is given by performing Monte Carlo integration to get corrected p values (searching over time). To do so, we use the fact that under the null hypothesis, T follows a multivariate t distribution with covariance matrix Σ^* and df degrees of freedom, where Σ^* is the covariance matrix of the time series of EWMA statistics Z .

Once $Var(z_t)$ is known, Σ^* can be calculated using the fact (Schmid & Schöne, 1997) that if $\{X_t\}$ is a weakly stationary process with mean θ_0 and autocovariance $\gamma(h)$ and $\gamma \in (0,1)$, then the off-diagonal terms of Σ^* are given by

$$Cov(z_t, z_{t-i}) = \frac{1}{2} \left((1 + \lambda)^i Var(z_{t-i}) + \frac{1}{(1 + \lambda)^i} (Var(z_t) - Var(z_i)) \right) \quad (8.10)$$

Knowing Σ^* allows us to accurately account for the temporal dependence (autocorrelation) in the time series in the Monte Carlo simulation.

Familywise error rate (FWER) control across the time series is provided by testing the observed T statistics against the maximum absolute T value obtained across the time series under the null hypothesis (Nichols & Holmes, 2002). We thus need to assess what are the distribution of maximum T values, where

$$T_{\max} = \max_{k \in [0, T]} \{ |T_k| \} \quad (8.11)$$

The Monte Carlo integration proceeds by generating a large number (we used 10,000 in all results presented here) n -length time series of multivariate random T values with covariance Σ^* . We use the multivariate T random number generator provided in Matlab (Mathworks, Natick, MA). $T_{\max}$ is saved from each time series, and the 95th percentile of the distribution of maxima provides a critical t value T^* for two-tailed FWER control.

Estimating change points

In the preceding section, we outlined a procedure for detecting *whether* a change in distribution has taken place at some point in the time series. In this section, we deal with the situation of estimating *when* exactly the change took place; that is, we want to estimate the unknown CP parameter τ . When a significant EWMA statistic indicates that a process is OOC (i.e., the state of activation has changed), it is fair to assume that the CP has occurred at an earlier time as changes are likely to continue for some time before the critical threshold for significance is reached.

Nishina (1992) proposed an approach for estimating the CP using the EWMA statistics directly. The last time point at which the process crosses θ_0 before exceeding the upper or lower control limit (and becoming significant) is the last *zero crossing*, and it is the estimate of τ . More formally, let us assume that N^{OOC} is the first time the EWMA statistic exceeds the

predefined threshold $c\sqrt{\text{Var}(z_t)}$, that is,

$$N^{\text{OOC}} = \min\{t \mid |z_t - \theta_0| > c\sqrt{\text{Var}(z_t)}\} \quad (8.12)$$

The CP estimator of τ for an *increase* in the process mean is given by the closest time point prior to N^{OOC} , which lies below θ_0 . That is, we define

$$\hat{\tau} = \max\{t \mid t \leq N^{\text{OOC}}, |z_t| \leq \theta_0\} \quad (8.13)$$

The CP estimator of τ for a *decrease* is defined in an analogous manner. This is the zero-crossing method, and its main advantages are that it is conceptually straightforward and is computationally efficient. A drawback is that it does not allow us to construct confidence intervals for τ . If we want these, we instead need to use a maximum likelihood estimate (MLE; see appendix B). In practice, we will be most interested in calculating the confidence set on the group level, and we discuss an approach toward obtaining such intervals using the zero-crossing method when discussing the group analysis.

Dealing with multiple change points

A limitation of the zero-crossing method is that only a single, initial CP is assessed, and there is no clear provision for assessing return to the baseline state or the presence of multiple CPs. One possible approach would be to assume a return to baseline once the EWMA statistic crosses back across the control limit. Hence, the total number of OOC points can give a rough indication of the activation duration, although durations are likely to be biased toward zero. Alternatively, we can use a Gaussian mixture model to classify each observation as either belonging to the baseline (in control) or activated (OOC) distribution. This procedure also allows us to estimate the length of time spent in the activated state.

In the mixture model approach, we model the fMRI time course as a mixture of two normal distributions, with different means, as follows: $X_0 \sim N(\theta_0, \sigma^2)$ for the baseline state, and $X_1 \sim N(\theta_1, \sigma^2)$ for the activated state (see Figure 8.1C for an example). We can write this mixture as

$$X = (1 - \Delta)X_0 + \Delta X_1 \quad (8.14)$$

where the random variable Δ is equal to one with probability p and equal to zero with probability $1 - p$. The density function of X can be written as

$$f_X(x) = (1 - p)f_{X_0}(x) + pf_{X_1}(x) \quad (8.15)$$

where $f_{x_i}(x)$ is the normal probability density function with mean θ_i and variance σ^2 .

We can fit this model to the data using maximum likelihood methods. The unknown parameters of the model are $(\theta_0, \theta_1, \sigma^2, p)$ and the log likelihood can be written:

$$l(\theta_0, \theta_1, \sigma^2, p | \mathbf{x}) = \sum_{i=1}^n \log(f_{x_i}(x_i)) \quad (8.16)$$

The parameters that maximize this term can be found using the Expectation-Maximization (EM) algorithm (see appendix C).

Once we have determined the MLEs of the parameters in this model, we need to classify each data point according to which state they belong. This can be done using Bayes' formula. The probability that a data point belongs in the active (OOC) state is given by

$$P(\text{active} | x_i) = \frac{pf_{x_1}(x_i)}{f_x(x_i)} \quad (8.17)$$

If $P(\text{active} | x_i) > 0.5$, then the time point is classified as belonging to the active state; otherwise, it is classified as belonging to the original baseline state.

HEWMA: A hierarchical extension of EWMA

The EWMA procedure outlined is suitable for studying a single time series for an individual subject. In fMRI analysis, we are typically most interested in detecting whether an effect is present over an entire group of subjects. Group changes are of primary interest in the neurosciences for making population inferences. For this reason, we developed a group analysis using a hierarchical (i.e., mixed-effects) model. Here, we introduce the HEWMA model that allows us to perform a mixed-effects analysis on fMRI group data using the same type of analysis that the EWMA method allowed for single-subject data. We use the EWMA statistic and covariance matrix defined previously to obtain the population average and its corresponding covariance matrix. Thereafter, the Monte Carlo procedure used in the EWMA framework is applied to get p values and test the hypothesis of consistent activation within the group.

Again, we are studying one voxel of the brain at a time (i.e., the massive univariate approach). We begin with the standard assumption that each subject's data have been warped to a standard brain atlas, or individual ROIs have been chosen, so that the time series from a brain area of interest

can be studied in a group. Suppose our study consists of m subjects. Let $\{x_t^i\}$ denote the data for subject i , where $i = 1, \dots, m$. Our hierarchical mixed-effects model takes the following form:

$$x_t^i = s_t^i + \varepsilon_t^i \quad (8.18)$$

$$s_t^i = s_t^{pop} + \eta_t^i \quad (8.19)$$

where η_t^i is a subject-level error term at time t , and s_t^i , the underlying signal for subject i , is either in a baseline or activated state at each time t according to

$$s_t^i = \begin{cases} \theta_0^i & \text{for } t = 1, \dots, \tau \\ \theta_1^i & \text{for } t = \tau + 1, \dots, N \end{cases} \quad (8.20)$$

Finally, the population signal

$$s_t^{pop} = \begin{cases} \theta_0^i & \text{for } t = 1, \dots, \tau \\ \theta_1^i & \text{for } t = \tau + 1, \dots, N \end{cases} \quad (8.21)$$

shows how the subject mean state values are drawn from a larger population. In matrix format, we can write Equation 8.18 as

$$X^i = S^i + E^i = I * S^i + E^i \quad (8.22)$$

where $X^i = (x_1^i, x_2^i, \dots, x_n^i)^T$, $S^i = (s_1^i, s_2^i, \dots, s_n^i)^T$, and $E^i = (\varepsilon_1^i, \varepsilon_2^i, \dots, \varepsilon_n^i)^T$. Here, $E^i \sim N(0, \Sigma_i)$, I is the $n \times n$ identity matrix, and the vector S^i can be considered the equivalent of beta weights — one weight for each time point. Similarly, we can write Equation 8.19 as

$$S^i = S_{pop} + U \quad (8.23)$$

where $S_{pop} = (s_1^{pop}, s_2^{pop}, \dots, s_n^{pop})^T$ and U is $N(0, \Sigma_B)$.

In the hierarchical setting, we can treat the between-subject variance in either of two ways. First, we could treat the vector S^i as beta weights and let U be an $n \times 1$ vector of iid $N(0, \sigma_B)$ random variables. Thus, the subject-level effect is allowed to vary over time. In this scenario, $S_B = I\sigma_B$, where σ_B is the variance of η_t . Alternatively, we could assume that S^i has an error term that is constant over time and let U be an $n \times 1$ vector of ones times a univariate $N(0, \sigma_B)$ random variable. We can write this as $U = \mathbf{1} * u$, where $\mathbf{1}$ is the vector of ones, and u is a univariate random variable. In this situation, $\Sigma_B = \mathbf{1}^T \mathbf{1} \sigma_B$.

However, in our implementation we use the first form of the covariance matrix.

Regardless, the model can be written in single-level format as

$$X^i = S_{pop} + U + E^i = S_{pop} + V^i \quad (8.24)$$

where $V^i \sim N(0, \Sigma_i + \Sigma_B)$; that is, the overall error term is distributed with a variance equal to the sum of between-subject and within-subject variances. Thus, in the HEWMA model the unknown parameters are, θ_0, θ_1, τ , and the variance components Σ_i and Σ_B . Note that if we have first performed a single-level analysis (EWMA) on each subject, we can use the results in the second level. In this case, we can assume that the within-subject variance components Σ_i are known, and brought forward, from the first level of analysis.

The EWMA statistic at time t for subject i , which we denote Z_t^i , is defined as in Equation 8.4. The HEWMA statistic, defined more formally below, is the weighted sum of the individual subjects' EWMA statistics. The subjects are weighted by the inverse of their respective variances, which are calculated as described in the EWMA section. However, in the group analysis, we will also need to take into consideration the between-subject variation. The total variation in an individual subject's EWMA statistic is written $V_i^* = \Sigma_i^* + \Sigma_B^*$, where Σ_i^* is the variance from the single-subject analysis, and Σ_B^* can be estimated using Equations 8.8 and 8.10 in combination with the fact that $\Sigma_i = I\sigma_B$. The only unknown term that needs to be estimated within the HEWMA framework is therefore the parameter σ_B . Hence, we can write $\Sigma_B^* = \alpha Q$, where Q denotes the known portion and α the unknown portion of the covariance matrix.

The HEWMA statistic and the parameter α are found iteratively in a manner outlined below.

Using an initial estimate of the entire relevant variance component, we can perform the weighted (hierarchical) regression. The HEWMA statistic is a weighted average of the individual EWMA statistics, with the weights inversely proportional to the total variance for each subject, that is,

$$Z_{pop} = \left[\sum_{i=1}^m W^i \right]^{-1} \sum_{i=1}^m W^i Z^i \quad (8.25)$$

where $W^i = V_i^{*-1}$, and the covariance matrix for Z_{pop} is given by

$$\left[\sum_{i=1}^m W^i \right]^{-1} \quad (8.26)$$

The estimation of the parameter α can be performed using a restricted maximum likelihood (ReML) approach (see appendix D), and the whole process of estimating Z_{pop} and its variance components is performed iteratively. This is equivalent to the variance component estimation procedure performed in SPM2 (Friston et al., 2002).

The final step in the HEWMA framework is performing a Monte Carlo simulation to get a corrected p value (searching over time). This is done as described in the EWMA section, except here we use Z_{pop} and its covariance matrix to calculate the relevant test statistics and define the multivariate t distribution. As a final note, we will often be interested in calculating a confidence interval for the CP on the group level. To do so, we can bootstrap the individual subjects' CP estimates, use this to estimate the distribution of $\hat{\tau}$, and thereby obtain confidence intervals.

Simulations

To test the EWMA and HEWMA methodology and the efficiency of the CP estimation procedure, we performed a set of three simulation studies outlined in detail below. In Simulation 1, we simulated a single-subject data set with four active regions, each with a different activation onset time, and estimated both the likelihood of activation and CPs using EWMA with the zero-crossing method. In Simulation 2, we assessed the FPR and power for the HEWMA method across values of the smoothing parameter λ . Simulation 3 assessed power and FPR across varying durations of baseline period activity.

Simulation 1

As shown in Figure 8.2, we constructed a 64×64 phantom image containing a square region of size 48×48 representing a human brain. The image intensities were assigned values of 1 or 0 for the points inside or outside the square, respectively. Four smaller squares, with dimensions 8×8 , were placed inside the larger square to simulate ROIs with static contrast to the larger square. To simulate a dynamic image series, this base image was re-created 250 times according to a boxcar paradigm consisting of a prolonged period of activation of length 50 time points during which the signal within the four squares increased to 2. The onset time of activation varied between the four regions and took values of 60, 80, 100, and 120 time points. Hence, each region had activation of similar length and intensity, but with varying onset times. Noise simulated using an autoregressive process of order 2 (AR(2)) model with standard deviation equal to 1 was added to each voxel's time course. We analyzed the simulated data set using EWMA with $\lambda = 0.2$ and an AR(2) noise model. We further estimated onset times (CPs) and p values for all active voxels.

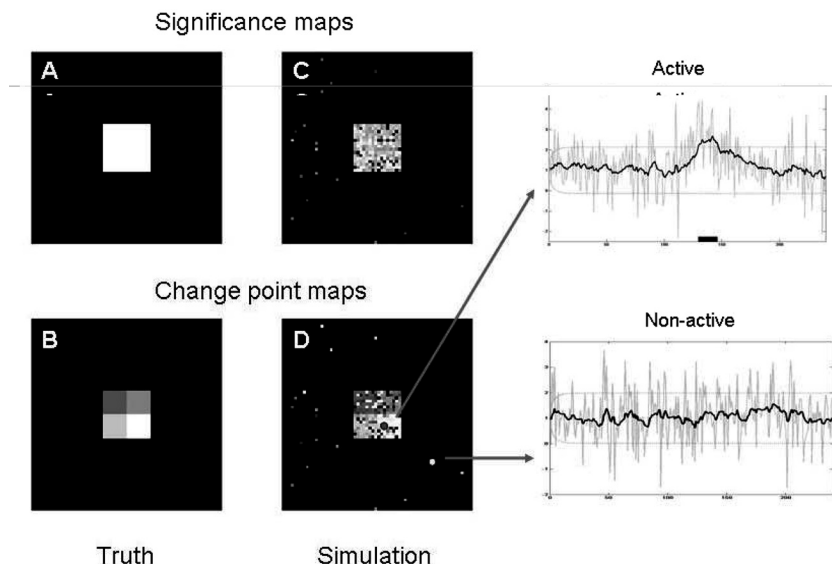


Figure 8.2 Results from a simulated single run experiment (Simulation 1) using an effect size equal to 1 and an AR(2) noise model. (A-B) The true significance and change point maps, (C) the significance map obtained from the EWMA analysis of the data (using $\lambda = 0.2$ and the AR(2) control bounds) and (D) the change point map estimated using the zero-crossing method. To the right are examples of both active and non-active time courses plotted together with their control limits. The data is represented by the light gray line and the EWMA statistic by the dark black line.

Simulation 2

The next simulation sought to study the FPR and perform power calculations for the HEWMA method. Actual fMRI noise was extracted from nonsignificant voxels of the brain (HEWMA detection $p > .95$), obtained from the experimental data described in the next section. In total, 50,350 noise time courses of length 215 time points were included in the study. The simulation mimicked a group analysis consisting of 20 subjects. For the FPR study, null hypothesis data with no activation were created by randomly sampling time series from the collection of noise time courses. This was done for each of 20 “subjects,” and a random between-subject variation with a standard deviation of size one third of the within-subject variation was added to each subject’s time course. A significance level of $\alpha = 0.05$ was used to determine “active” voxels. For the power calculations, the same procedure was repeated, with the difference that an active

period of length 50 time points was added to the noise data, with intensity equivalent to a Cohen's d of 0.5. This coincides with values observed in experimental data (Wager, Vazquez, Hernandez, & Noll, 2005).

The HEWMA method was performed on 5,000 replications of each of these two data types for λ values set equal to 0.1, 0.3, 0.5, 0.7, and 0.9. The analysis was further performed using error models in the HEWMA framework corresponding to WN and AR(1), AR(2), and ARMA(1,1) noise. For each simulation, the first 60 time points were used as a baseline period.

Simulation 3

The procedure was identical to that for Simulation 2, except that in this simulation, the baseline length was 20, 40, 60, or 80 time points. The analysis was again performed using error models in the HEWMA framework corresponding to WN and AR(1), AR(2), and ARMA(1,1) noise. For each simulation, λ was set equal to 0.2, and a significance level of $\alpha = 0.05$ was used.

Experimental fMRI data collection and analysis

Participants

We applied the HEWMA method to data from 30 participants scanned with BOLD fMRI at 3 T (GE, Milwaukee, WI). The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the University of Michigan institutional review board. Six participants were excluded because of motion or nonlinear normalization artifacts, leaving 24 participants.

Task design

The task used was a variant of a well-studied laboratory paradigm for eliciting anxiety (Dickerson & Kemeny, 2004; Gruenewald, Kemeny, Aziz, & Fahey, 2004; Roy, Kirschbaum, & Steptoe, 2001), shown in Figure 8.3.

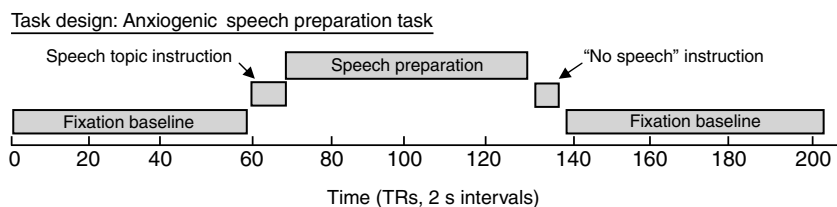


Figure 8.3 Task Design: Anxiogenic speech preparation task. Here TR denotes the time resolution of the study.

The design was an off-on-off design, with an anxiety-provoking speech preparation task occurring between lower-anxiety resting periods. Participants were informed that they were to be given 2 min to prepare a 7-min speech, and that the topic would be revealed to them during scanning. They were told that after the scanning session, they would deliver the speech to a panel of expert judges, although there was “a small chance” that they would be randomly selected not to give the speech.

After the start of fMRI acquisition, participants viewed a fixation cross for 2 min (resting baseline). At the end of this period, participants viewed an instruction slide for 15 sec that described the speech topic, which was to speak about “why you are a good friend.” The slide instructed participants to be sure to prepare enough for the entire 7-min period. After 2 min of silent preparation, another instruction screen appeared (a “relief” instruction, 15-sec duration) that informed participants that they would not have to give the speech. An additional 2-min period of resting baseline followed, which completed the functional run.

Heart rate was monitored continuously, and heart rate increased after the topic presentation, remained high during preparation, and decreased after the relief instruction (data are presented elsewhere). Because this task involved a single change in state, as in some previous fMRI experiments (Breiter & Rosen, 1999; Eisenberger, Lieberman, & Williams, 2003), and the precise onset time and time course of subjective anxiety were unknown, this design was a good candidate for the HEWMA analysis.

Image acquisition

A series of 215 images was acquired using a T2*-weighted, single-shot reverse spiral acquisition (gradient echo, TR = 2000, TE = 30, flip angle = 90°) with 40 sequential axial slices (FOV = 20, 3.12 × 3.12 × 3 mm, skip 0, 64 × 64 matrix). This sequence was designed to enable good signal recovery in areas of high susceptibility artifact (e.g., orbitofrontal cortex, OFC). High-resolution T1 spoiled gradient recall (SPGR) images were acquired for anatomical localization and warping to standard space.

Image analysis

Offline image reconstruction included correction for distortions caused by magnetic field inhomogeneity. Images were corrected for slice acquisition timing differences using a custom four-point sinc interpolation and realigned (motion corrected) to the first image using automated image registration (AIR; Woods, Grafton, Holmes, Cherry, & Mazziotta, 1998). SPGR images were coregistered to the first functional image using a

mutual information metric (SPM2). When necessary, the starting point for the automated registration was manually adjusted and rerun until a satisfactory result was obtained. The SPGR images were normalized to the Montreal Neurological Institute (MNI) single-subject T1 template using SPM2 (with the default basis set). The warping parameters were applied to functional images, which were then smoothed with a 9-mm isotropic Gaussian kernel.

Individual-subject data were subjected to linear detrending across the entire session (215 images) and analyzed with EWMA. An AR(2) model was used to calculate the EWMA statistic (Z) and its variance, and Z and variance estimates were carried forward to the group-level HEWMA analysis. We used custom software (see acknowledgment section for download information and appendices) to calculate statistical maps throughout the brain, including HEWMA (group) t and p values for activations (increases from baseline) and deactivations (decreases from baseline); individual and group CPs (calculated on the group HEWMA t time series) using the zero-crossing method, activation duration as estimated by the number of OOC points, and CP and run-length estimates using the Gaussian mixture model described above.

Significant voxels were classified into sets of voxels showing similar behavior using k -means clustering. We considered activations and deactivations separately, and used k -means clustering on the group CP and longest activated run length (from the mixture model) to assign voxels into classes with similar behavior. To do this, we used the k -means algorithm implemented in Matlab 7.4 with the $v \times 2$ matrix of values for the v significant activated (or deactivated) voxels as input. The number of classes was determined by visual inspection of the joint histogram of CP and duration values. Twelve classes were used for the analyses reported here. Sets of contiguous suprathreshold voxels (*regions*) of the same class were the unit of analysis for interpretation.

Examining the systematic features in the time courses of ROIs permitted us to make inferences about the role the region played in the speech preparation task. Rather than limitation to testing whether a voxel is activated during the preparation interval on average compared with baseline periods, the HEWMA method can detect a number of different types of interesting systematic features, including sustained activation during the preparation task, transient activation during instruction presentation, fluctuations in baseline activity that may be related to the start of scanning, and others. Of particular interest are voxels with activation onset that is near the time of task onset (at 60 TRs or 120 sec) and with activity that is sustained throughout the task (at least 60 TRs/120 sec), which may reflect sustained anxiety.

Results

Simulation results

Simulation 1

We used EWMA to create a significance map and a CP map (CPM) that accurately depict the difference in onset time between regions. Setting the smoothing parameter $\lambda = 0.02$ and the significance level $\alpha = 0.05$, we analyzed each voxel using the EWMA procedure outlined in the Methods section. Figure 8.2A depicts the theoretical significance map, with equal amount of activation present within each of the four active regions. Figure 8.2C depicts the actual significance map obtained using EWMA. This indicates that we were able to accurately detect a large number of active voxels within the region of activity, with a minimal number of false positives outside the region. Figure 8.2B depicts the theoretical CPM, in which the intensity varies depending on the onset time. Figure 8.2D depicts the actual CPM obtained by calculating the zero-crossing CP estimate for each voxel deemed active in the prior analysis. The CPM provided accurate estimates of the onset times for the four regions, as indicated by the similar intensity values for the true values (Figure 8.2C) and estimates (Figure 8.2D). Examination of the error in CP estimation showed a distribution that was centered at zero with a slight left skew (mean = -2.0 , SD [standard deviation] = 6.3 , median = 0 , and IQR (interquartile range) = 5).

Simulation 2

Figure 8.4A and 8.4B shows the FPR and power calculations for each noise type as a function of the smoothing parameter λ . Figure 8.4A shows that the number of false positives increased for each noise type as a function of λ (i.e., with less smoothing). This is natural as low values of λ entail a greater amount of smoothing and minimize the risk of the null hypothesis data venturing too far from the baseline mean. The nominal α level of $.05$ is shown by the horizontal dashed line. As λ increased, the amount of smoothing decreased, and the FPR exceeded $\alpha = 0.05$. The ARMA model performed worse than the other models in FPR and power, although the other models appeared to behave in a similar manner (but they are somewhat conservative, with FPRs near 0.01 with low λ). All models controlled the FPR appropriately with $\lambda \leq .4$.

Studying Figure 8.4B, it appears that although the power increased slightly for each noise type as a function of λ , it did not vary in a significant manner. All noise models gave roughly equivalent results. In summary, our simulation studies indicated that a low value of λ will give a test with high power and low FPRs. Increasing the value of λ will lead

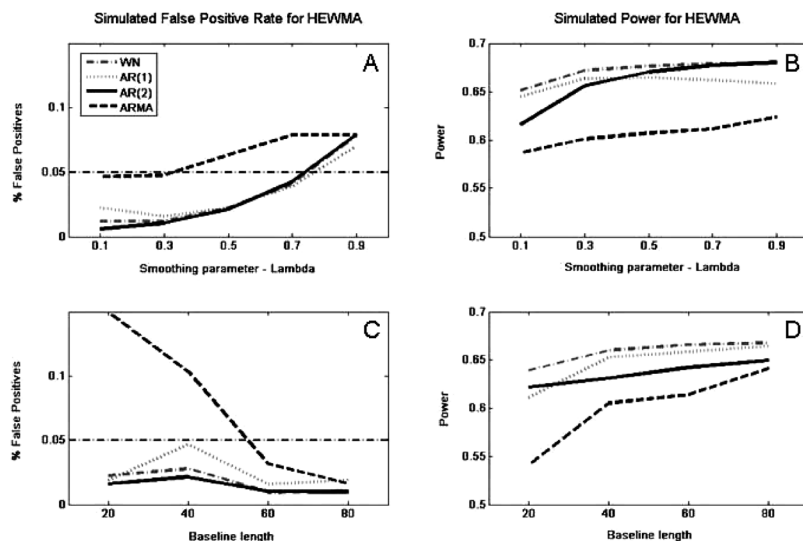


Figure 8.4 (A-B) Simulation 2. Simulated power and false positive rates for HEWMA with varying smoothness parameter λ , and fixed baseline length of 60 TRs. (C-D) Simulation 3. Same plots with fixed $\lambda = 0.2$ and varying baseline length.

to a slight increase in power, but at the cost of an increase in FPR. In the continuation, we used $\lambda = 0.2$.

Simulation 3

Figure 8.4C and 8.4D shows the FPR and power calculations for each noise type as a function of the baseline period length. Studying Figure 8.4C, it is clear that the number of false positives decreased for each noise type as the baseline period increased. This is natural as a long baseline period allows for more data to accurately estimate the parameters of the model. The ARMA model performed significantly worse than the other models and gave rise to a dramatically inflated FPR for baseline lengths less than 60 time points. This performance may be due to the fact that the baseline period was too short to get an accurate estimate of the variance components for this model type as the ARMA parameter estimation was more complex than for the other models (i.e., MLE vs. method of moments). Studying Figure 8.4D, it appears that although the power increased slightly for each noise type as a function of baseline length, it did not vary substantially across baseline durations. Again, the ARMA model performed worse than the other models.

In summary, our simulation studies indicated that an increased baseline leads to increased power as well as decreased FPR, which is advantageous. However, it is interesting to note that the WN, AR(1), and AR(2) models were all robust enough to handle a baseline period as short as 20 time points. Naturally, these values may depend to some degree on the noise characteristics. The ARMA model, on the other hand, required a baseline of at least 60 time points.

fMRI results

As expected, compared with baseline, the task elicited both significant activity increases (red in Figure 8.5A) and decreases (blue in Figure 8.5A) in a number of brain regions. Increases in visual cortex, superior parietal cortex, temporal cortex, and lateral and medial prefrontal cortices are shown in red. The activation was consistent with what might be expected in a cognitively complex task, which involved visual cues at two periods during the task, the mental effort and subvocal rehearsal required to prepare a speech, and the anxiety elicited by the task context. In particular, the rostral PFC has been strongly implicated in processing of self-relevant information and the representation and regulation of aversive emotional states (Ochsner et al., 2004; Phan, Wager, Taylor, & Liberzon, 2002; Quirk & Gehlert, 2003; Quirk, Russo, Barron, & Lebron, 2000; Ray et al., 2005; Wager et al., 2003). Information about the onset and duration of activation provided by the HEWMA analysis can constrain interpretation of the roles of these regions in task performance.

Figure 8.5B shows CPMs of the time of onset of activity (indicated by color) for activated regions. Figure 8.5C shows a histogram of CP values, with onsets color-coded according to onset time in Figure 8.5B. The range of estimated onset times, from around 15 TRs (30 sec) to around 180 TRs from the start of scanning indicates that different regions were activated at very different times during scanning. Most significant voxels showed activation onsets around the time of task onset (60 TRs), when the visual cue to begin speech preparation was presented. These include rostral medial prefrontal cortex (rMPFC), lateral OFC, occipital and parietal cortices (yellow in Figure 8.5B). However, other regions became active before the instructional cue, signaling that they may play a role in anticipation or preparation to begin the task. These regions include some visual areas, superior temporal cortex, and subgenual anterior cingulate. Another set of regions appeared to be activated before task onset but showed decreases during speech preparation; these include the amygdala and extended amygdala, ventral anterior insula, and striatum.

Likewise, activation duration estimates from the Gaussian mixture model ranged from transient increases (approximately 13 TRs or 26 sec) to

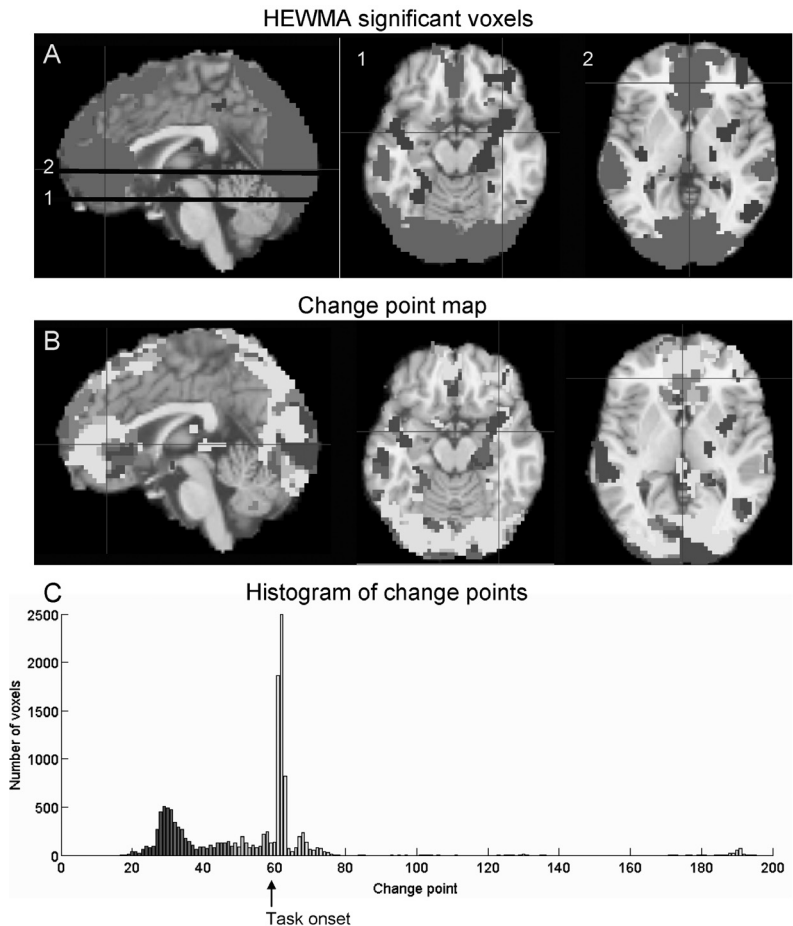


Figure 8.5 (A) A sagittal slice (left) and axial slices (center and right) showing significantly activated voxels in red and deactivated voxels in blue in the speech preparation task. The threshold used was $p < 0.05$, corrected for search over time (but not space). (B) Group change-point (CP) estimates using the zero-crossing method. Earlier CPs are shown in red, and later ones in yellow. (C) A histogram of CPs for significant voxels, colored as in (B). Most voxels responded around the time of task onset (yellow), although another set of voxels showed systematic activation before the task onset (red).

sustained increases (~170 TRs, 340 sec). The two-dimensional histogram of significant voxels by CP and estimated duration is shown in Figure 8.6A. The diversity of onset times and durations suggests that a variety of different GLM models would be required to detect these activations.

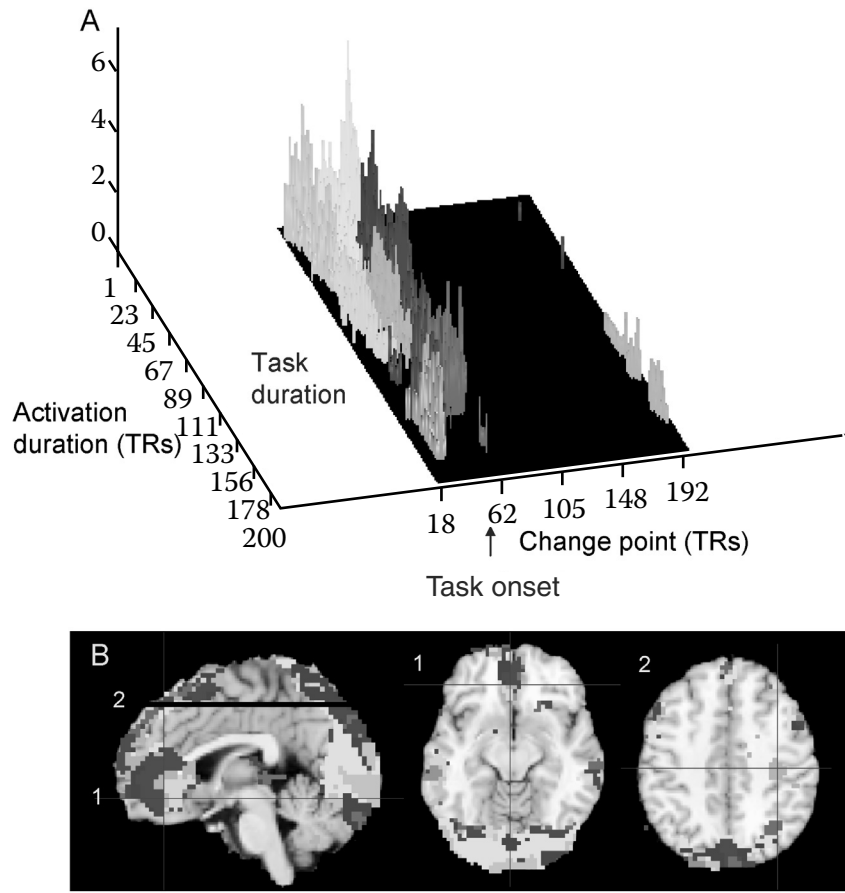


Figure 8.6 (A) A 2-D histogram of significant voxels by CP (x-axis) and activation duration (y-axis), estimated using the mixture model. Log frequency is shown on the z-axis. Bars are color-coded according to class based on k-means classification, Voxels of the same color thus have similar CP and duration profiles. (B) A sagittal slice (left) and axial slices (right) showing significant voxels, color-coded according to class as in (A). The slices show adjacent regions of rostral medial prefrontal cortex that are active throughout the task and begin to respond upon task presentation (red) or in anticipation (cyan). A sensorimotor region (green) has a much later onset of activity.

The *k*-means classification provides a way to identify classes of activated voxels with similar onset and duration estimates. Class membership is indicated by color in the histogram in Figure 8.6A, and Table 8.1 lists the number of voxels in each of the 12 classes specified, along with means for CPs and duration estimates and their standard deviations across

Table 8.1 Classes of Significant Voxels from K-Means Clustering

Num. Voxels	CP	Std	Duration	Std	Color
2778	61.3	1.5	13.1	4.48	yellow
1368	61.1	1.8	81.5	10.74	red
1290	29.5	4.3	23.1	13.21	light green
1237	61.3	2.2	43.2	7.86	blue
818	59.9	4.3	142.4	14.37	magenta
773	35.7	6.2	151.6	8.99	orange
614	31.6	3.9	79.4	14.74	cyan
613	49.7	3.6	79.0	13.26	green
594	29.6	2.0	173.8	8.17	red
482	46.5	3.3	22.4	12.61	cyan
208	189.6	5.1	149.8	21.9	green
30	170.2	23.9	42.4	25.26	blue

Note: Classification of significant voxels by change point (CP) and maximum-likelihood duration of the longest activated period using k-means clustering. A twelve-class solution was chosen based on visual inspection of the histogram (Fig. 8.6A). Colors correspond to those in Fig. 8.6.

voxels within the class. The largest class of activated voxels showed both onset times around the start of the task and durations lasting as long (60 TRs) or slightly longer than the task. The red and blue classes (Table 8.1, Figures 8.6 and 8.7), for example, became active at task onset. Activity in the red class, which includes rMPFC, persisted for 80 TRs (Figure 8.6B, 40 sec past the onset of the relief cue), whereas activity in the blue class, which includes precuneus, diminished before the relief cue (43 TRs/86-sec duration on average, compared to 120 sec of speech preparation). These regions may be related to the anxiety inherent in task performance.

Other classes seem clearly related to visual processing of the cues. The yellow and light green classes were activated at task onset or before (ramping up before the instruction) and persisted for only approximately 25 to 45 sec after the onset of the 15-sec visual instruction. Yet other classes, such as the green class that began to be active around 150 TRs (300 sec), after the relief cue, have more ambiguous interpretations. These regions may be related to relief, or they may show a drift in signal at the end of the session due to fatigue or increased movement.

Examining the group time courses of voxels in these regions corroborates this view and provides additional evidence. The sensorimotor region that comprises most of the late-onset green class is shown in Figure 8.7A, and it appears to drift upward at the end of scanning. Notably, the deviation is in the same direction for all participants, so is unlikely to be related to head movement unless participants moved in a systematic way at the end

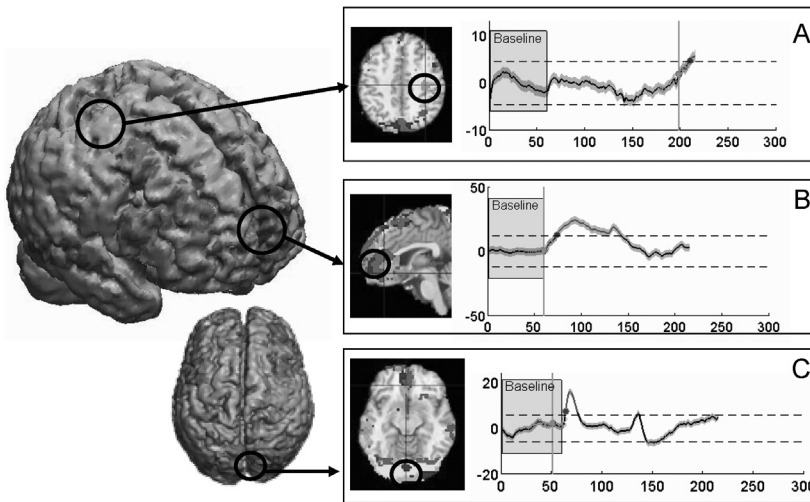


Figure 8.7 The brain surface is shown in lateral oblique and axial views. Significant voxels are colored according to significance. Increases are shown in red-yellow, and decreases are shown in light-dark blue. (A) A time course from a region showing late-onset responses (green class in Fig. 6). The time course was extracted from a single representative voxel in the region. The baseline period is indicated by the shaded gray box, and the HEWMA-statistic is shown by the thick black line ($\pm$ one standard error across participants, shown by gray shading). The control limits are shown by dashed lines. This activation is unlikely to be anxiety-related. (B) A similar plot for the rMPFC (red class). The time course parallels the structure of the task in that activation begins around the onset of speech instructions and is sustained throughout the duration of the preparation interval. (C) HEWMA time series plot for a representative region in the primary/secondary visual cortex (yellow class). Activity peaks twice, during each presentation of visual instructions.

of scanning, but may be related to other nuisance variables. rMPFC voxels, shown in Figure 8.7B, closely parallel the time course of subjective anxiety and heart rate increases in the task. Occipital voxels, such as the one shown in Figure 8.7C (yellow class), became active when the instruction cue was presented and again when the relief cue was presented, consistent with a large literature implicating these regions in visual processing. The baseline period in Figure 8.7 is indicated by the shaded gray box in each panel, and the HEWMA statistic time course is shown by the thick black line ($\pm$ one standard error across participants, shown by gray shading). Control limits are shown by dashed lines; thus, the region is significantly activated if the HEWMA statistic exceeds the control limit at any time point.

Importantly, the time course of activation in rMPFC parallels the structure of the task in that activation began around the onset of speech instructions (CP estimate 122 sec, actual task onset 120 sec) and was sustained throughout the duration of the preparation interval. This region has been related in many neuroimaging studies to subjective anxiety and self-referential processing (Breiter & Rosen, 1999; Dougherty et al., 2004; Eisenberger et al., 2003; Ray et al., 2005; Wang et al., 2005) and is thus a neurophysiologically plausible candidate to show sustained activation related to the anxiogenic task.

Discussion

Typically, statistical methods in fMRI can be categorized into two broad categories: hypothesis and data-driven approaches. Hypothesis-driven approaches test whether activity in a brain region is systematically related to some known input function. In this approach, typically the GLM is used to test for differences in activity among psychological conditions or groups of participants. However, for many psychological processes, the precise timing and duration of psychological activity can be difficult to specify in advance. In this situation, the GLM approach becomes impractical as the psychological activity cannot be specified a priori. Data-driven methods, such as ICA, give an account of the data using few a priori assumptions. Instead, they attempt to characterize reliable patterns in the data and relate those patterns to psychological activity post hoc. The main drawback of these methods is that they do not provide statistics for making inferences about whether a component varies over time and when changes occur in the time series.

The purpose of the HEWMA method is to allow for the detection of systematic changes in activity with a variety of onset times and activation durations. The differences in onset and duration are likely to reflect differences in functional anatomy, that is, the way in which each activated region participates in the task. Both the EWMA and HEWMA methods for fMRI data analysis are designed to detect regions of the brain where the signal deviates reliably from a baseline state. The methods make no a priori assumption about the behavior of these changes, and the method will detect activation and deactivation as well as regions with both short and prolonged activation duration. In this sense, both EWMA and HEWMA can be thought of as searches for activity differences across time, correcting for the multiple comparisons tested and accounting for the correlation among observations.

Once a systematic deviation from baseline has been detected, the second step in the analysis entails estimating *when* exactly the change took place as well as the recovery time (if any). This estimation procedure can

be performed using the zero-crossing method, the MLE, or a Gaussian mixture model. Once these estimates are obtained, we can cluster the active voxels into groups with estimates that behave in a similar manner. This allows us to classify regions and even discard regions for which the activation was triggered by effects such as drift or movement. As a final step, after estimating the CP for each active voxel in the brain, we summarize the results in a CPM. A CPM is an image of the brain with a color-coded CP mask superimposed, with its intensity varying depending on the estimated onset time of activation.

The main decisions that need to be made before applying these methods to experimental data are the choice of the smoothing parameter λ and the length of the baseline period. According to our power and FPR analyses, a relatively low value of λ gives high power, with a strong control of the FPR. In our analysis of experimental data, we typically choose a value of 0.2 as this appears to give rise to an adequate amount of smoothing for the analysis of fMRI data. The optimal value of λ will vary depending on whether brief or sustained changes are of greater interest, with lower values more appropriate for more sustained activity. The length of the baseline period is another issue. The data within this period are used to estimate the baseline mean as well as the within-subject variation, and our simulations showed that, unsurprisingly, longer baseline periods produced more accurate estimates. However, our analysis indicated that the method is relatively stable for even very short baseline periods for WN, AR(1), and AR(2) models. The ARMA model showed increased sensitivity to short baseline periods with drastically increased FPR and decreased power, and a longer baseline period (>60 time points) is recommended. The AR(2) model may be most appropriate for fMRI data as it has the flexibility to model periodic noise oscillations that are often produced in fMRI as a result of physiological changes (e.g., pulsatile motion of the brain due to breathing and cardiac activity).

HEWMA appears to be an appropriate analysis for group fMRI data, particularly when it is not possible to replicate experimental manipulations within subjects (e.g., a state anxiety induction that cannot be repeated without changing the psychological nature of the state). Emotional responses are one prime candidate for applications of the method. But, there are a number of other domains in which it may be useful as well, and the method applies to any longitudinal data with enough observations so that repeated-measures analysis of variance (ANOVA) (for example) is impractical. HEWMA may be particularly useful for arterial spin labeling and perfusion MRI studies, which measure brain activity over time without the complicating factors of signal drift and highly colored noise in fMRI (Liu, Wong, Frank, & Buxton, 2002; Wang et al., 2005).

Another potential use is in identifying voxels of interest and characterizing brain responses in “ecologically valid” tasks, such as free viewing of films. For example, according to one article, participants watched a 60-min segment of an action movie (Hasson, Nir, Levy, Fuhrmann, & Malach, 2004). The investigators examined the time course of activity throughout the brain and assessed whether increases in particular regions were systematically related to features of the film (e.g., presentation of scenes, hands, faces). HEWMA could be used in this situation to identify voxels that responded consistently across participants during viewing, which would reduce false positives by providing a reduced set of voxels of interest and providing some quantitative tools for characterizing the duration and number of activated periods.

HEWMA could also be applied to changes in state-related activity evoked by learning (e.g., tonic increases brain activity as a function of expertise) or to studies of tonic increases following solutions to “insight” problem-solving tasks. Expertise results in functional and structural reorganization of cortex (Kilgard & Merzenich, 1998; Kourtzi, Betts, Sarkheil, & Welchman, 2005), and HEWMA could be used to more precisely characterize the time course of both types of changes (i.e., do shifts occur gradually or suddenly?). In an insight task, participants are presented with a problem that requires a novel combination of elements (Bowden, Jung-Beeman, Fleck, & Kounios, 2005; MacGregor, Ormerod, & Chronicle, 2001). Once participants solve the task, there is a qualitative shift in their understanding of how the elements of the problem relate that cannot be reversed (the solution is “obvious” once one knows it). This profound shift is poorly understood, in part because appropriate methods have not been devised to study its brain mechanisms in healthy participants.

Another potential use is in longitudinal studies of brain function or structure, and how they change with development or with the progression of a neurological or psychiatric disorder. For example, Mayberg and colleagues conducted several longitudinal studies of resting fluorodeoxyglucose (FDG) PET activity in depressed patients over the course of treatment (Goldapple et al., 2004; Mayberg et al., 2002). The time course of brain activity changes throughout the treatment process are unknown, and HEWMA could be used to locate regions that respond to treatment and identify the time at which they do so.

Conclusion

In this chapter, we developed a hybrid data and model-driven approach, HEWMA, that can be used to make inferences about individual or group fMRI activity, even when conditions are not replicated (e.g., a single

experimental induction of emotion). This approach can complement GLM-based and purely data-driven methods (such as ICA) by providing inferences about whether, when, and for how long systematic state-related activation occurs in a particular brain region. Although it is developed here for fMRI data analysis, the method could be useful in detection of deviation from a baseline state in any type of time series data, including arterial spin labeling (ASL), longitudinal studies of brain structure, or positron emission tomographic (PET) activity, and others.

The HEWMA method is an extension of EWMA, a time series analysis method in statistical process control theory and CP theory, to multisubject data. It permits population inference and can be used to analyze fMRI data voxelwise throughout the brain, data from ROIs, or temporal components extracted using ICA or similar methods. Simulations showed that the method has acceptable FPR control, and application to an fMRI study of anxiety showed that it produces reasonable and novel results with empirical data. A toolbox implementing all functions in Matlab is freely available from the authors (see next section).

Acknowledgment

We would like to thank Christian Waugh, Doug Noll, and Luis Hernandez for providing data with which to develop the model. Software implementing the EWMA and HEWMA analyses is available for download from <http://www.columbia.edu/cu/psychology/tor/> or by contacting the authors.

References

- Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. II. Spatially smoothed data collected under null-hypothesis and experimental conditions. *Neuroimage*, 5, 199–212.
- Beckmann, C. F., & Smith, S. M. (2004). Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23, 137–152.
- Beckmann, C. F., & Smith, S. M. (2005). Tensorial extensions of independent component analysis for multisubject FMRI analysis. *Neuroimage*, 25, 294–311.
- Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences*, 9, 322–328.
- Breiter, H. C., & Rosen, B. R. (1999). Functional magnetic resonance imaging of brain reward circuitry in the human. *Annals of the New York Academy of Sciences*, 877, 523–547.
- Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (2nd ed.). New York: Springer-Verlag.

- Calhoun, V. D., Adali, T., & Pekar, J. J. (2004). A method for comparing group fMRI data using independent component analysis: Application to visual, motor and visuomotor tasks. *Magnetic Resonance Imaging*, 22, 1181–1191.
- Calhoun, V. D., Adali, T., Stevens, M. C., Kiehl, K. A., & Pekar, J. J. (2005). Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *Neuroimage*, 25, 527–538.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J. K., Holyoak, K. J., et al. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *Neuroimage*, 14, 1136–1149.
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130, 355–391.
- Dougherty, D. D., Rauch, S. L., Deckersbach, T., Marci, C., Loh, R., Shin, L. M., et al. (2004). Ventromedial prefrontal cortex and amygdala dysfunction during an anger induction positron emission tomography study in patients with major depressive disorder with anger attacks. *Archives of General Psychiatry*, 61, 795–804.
- Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? An FMRI study of social exclusion. *Science*, 302, 290–292.
- Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crises? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11th, 2001. *Journal of Personality and Social Psychology*, 84, 365–376.
- Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: Applications. *Neuroimage*, 16, 484–512.
- Goldapple, K., Segal, Z., Garson, C., Lau, M., Bieling, P., Kennedy, S., et al. (2004). Modulation of cortical-limbic pathways in major depression: Treatment-specific effects of cognitive behavior therapy. *Archives of General Psychiatry*, 61, 34–41.
- Gruenewald, T. L., Kemeny, M. E., Aziz, N., & Fahey, J. L. (2004). Acute threat to the social self: shame, social self-esteem, and cortisol activity. *Psychosomatic Medicine*, 66, 915–924.
- Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., & Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biological Psychiatry*, 53, 494–501.
- Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. *Science*, 303, 1634–1640.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, 57, 1–17.
- Kilgard, M. P., & Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science*, 279, 1714–1718.
- Kourtzi, Z., Betts, L. R., Sarkheil, P., & Welchman, A. E. (2005). Distributed neural plasticity for shape learning in the human visual cortex. *PLoS Biology*, 3, e204.

- Koyama, T., McHaffie, J. G., Laurienti, P. J., & Coghill, R. C. (2005). The subjective experience of pain: Where expectations become reality. *Proceedings of the National Academy of Sciences*, 102, 12950–12955.
- Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., et al. (1998). Emotional arousal and activation of the visual cortex: an fMRI analysis. *Psychophysiology*, 35, 199–210.
- Liu, T. T., Wong, E. C., Frank, L. R., & Buxton, R. B. (2002). Analysis and design of perfusion-based event-related fMRI experiments. *Neuroimage*, 16, 269–282.
- Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements. *Technometrics*, 32, 1–29.
- MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 176–201.
- Mayberg, H. S., Silva, J. A., Brannan, S. K., Tekell, J. L., Mahurin, R. K., McGinnis, S., et al. (2002). The functional neuroanatomy of the placebo effect. *American Journal of Psychiatry*, 159, 728–737.
- McIntosh, A. R., Chau, W. K., & Protzner, A. B. (2004). Spatiotemporal analysis of event-related fMRI data using partial least squares. *Neuroimage*, 23, 764–775.
- McKeown, M. J., & Sejnowski, T. J. (1998). Independent component analysis of fMRI data: examining the assumptions. *Human Brain Mapping*, 6, 368–372.
- Montgomery, D. C. (2000). *Introduction to statistical quality control* (4th ed.). New York: Wiley.
- Neubauer, A. S. (1997). The EWMA control chart: properties and comparison with other quality-control procedures by computer simulation. *Clinical Chemistry*, 43, 594–601.
- Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15, 1–25.
- Nishina, K. (1992). A comparison of control charts from the viewpoint of change-point estimation. Quality and reliability. *Engineering International*, 8, 537–541.
- Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., et al. (2004). For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage*, 23, 483–499.
- Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage*, 16, 331–348.
- Pignatiello, J. J., & Samuel, T. R. (2001). Estimation of the change point of a normal process mean in SPC applications. *Journal of Quality Technology*, 33, 82–95.
- Quirk, G. J., & Gehlert, D. R. (2003). Inhibition of the amygdala: Key to pathological states? *Annals of the New York Academy of Science*, 985, 263–272.

- Quirk, G. J., Russo, G. K., Barron, J. L., & Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *Journal of Neuroscience*, 20, 6225–6231.
- Ray, R. D., Ochsner, K. N., Cooper, J. C., Robertson, E. R., Gabrieli, J. D., & Gross, J. J. (2005). Individual differences in trait rumination and the neural systems supporting cognitive reappraisal. *Cognitive, Affective, and Behavioral Neuroscience*, 5, 156–168.
- Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1, 239–250.
- Roy, M. P., Kirschbaum, C., & Steptoe, A. (2001). Psychological, cardiovascular, and metabolic correlates of individual differences in cortisol stress recovery in young men. *Psychoneuroendocrinology*, 26, 375–391.
- Schmid, W. (1997). *On EWMA charts for time series* (Vol. 5). Heidelberg: Physica-Verlag.
- Schmid, W., & Schöne, A. (1997). Some properties of the EWMA control chart in the presence of autocorrelation. *Annals of Statistics*, 25, 1277–1283.
- Shehab, R. L., & Schlegel, R. E. (2000). Applying quality control charts to the analysis of single-subject data sequences. *Human Factors*, 42, 604–616.
- Siegmund, D. (1986). Boundary crossing probabilities and statistical applications. *Annals of Statistics*, 14, 361–404.
- Taylor, S. F., Phan, K. L., Decker, L. R., & Liberzon, I. (2003). Subjective rating of emotionally salient stimuli modulates neural activity. *Neuroimage*, 18, 650–659.
- Tugade, M. M., & Fredrickson, B. L. (2004). Resilient individuals use positive emotions to bounce back from negative emotional experiences. *Journal of Personality and Social Psychology*, 86, 320–333.
- Wager, T. D. (2005). The neural bases of placebo effects in pain. *Current Directions in Psychological Science*, 14, 175–179.
- Wager, T. D., Hernandez, L., Jonides, J., & Lindquist, M. (2007). Elements of functional neuroimaging. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *The handbook of psychophysiology* (3rd ed.). New York: Cambridge University Press 1, 19–55.
- Wager, T. D., Jonides, J., Smith, E. E., & Nichols, T. E. (2005). Towards a taxonomy of attention-shifting: Individual differences in fMRI during multiple shift types. *Cognitive, Affective, and Behavioral Neuroscience*, 5, 127–143.
- Wager, T. D., Phan, K. L., Liberzon, I., & Taylor, S. F. (2003). Valence, gender, and lateralization of functional brain anatomy in emotion: A meta-analysis of findings from neuroimaging. *Neuroimage*, 19, 513–531.
- Wager, T. D., Vazquez, A., Hernandez, L., & Noll, D. C. (2005). Accounting for non-linear BOLD effects in fMRI: parameter estimates and a model for prediction in rapid event-related studies. *Neuroimage*, 25, 206–218.
- Wang, J., Rao, H., Wetmore, G. S., Furlan, P. M., Korczykowski, M., Dinges, D. F., et al. (2005). Perfusion functional MRI reveals cerebral blood flow pattern under psychological stress. *Proceedings of the National Academy of Science of the United States of America*, 102, 17804–17809.

- Woods, R. P., Grafton, S. T., Holmes, C. J., Cherry, S. R., & Mazziotta, J. C. (1998). Automated image registration: I. General methods and intrasubject, intramodality validation. *Journal of Computer Assisted Tomography*, 22, 139–152.
- Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited — again. *Neuroimage*, 2, 173–181.
- Yantis, S., Schwarzbach, J., Serences, J. T., Carlson, R. L., Steinmetz, M. A., Pekar, J. J., et al. (2002). Transient neural activity in human parietal cortex during spatial attention shifts. *Nature Neurosci*, 5, 995–1002.