Hierarchical Mixture Models in Neurological Transmission Analysis

Mike West

Journal of the American Statistical Association, Vol. 92, No. 438. (Jun., 1997), pp. 587-606.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199706%2992%3A438%3C587%3AHMMINT%3E2.0.CO%3B2-X

Journal of the American Statistical Association is currently published by American Statistical Association.

Hierarchical Mixture Models in Neurological Transmission Analysis

Mike WEST

Hierarchically structured mixture models are studied in the context of data analysis and inference on neural synaptic transmission characteristics in mammalian, and other, central nervous systems. Mixture structures arise due to uncertainties about the stochastic mechanisms governing the responses to electrochemical stimulation of individual neurotransmitter release sites at nerve junctions. Models attempt to capture such scientific features as the sensitivity of individual synaptic transmission sites to electrochemical stimuli and the extent of their electrochemical responses when stimulated. This is done via suitably structured classes of prior distributions for parameters describing these features. Such priors may be structured to permit assessment of currently topical scientific hypotheses about fundamental neural function. Posterior analysis is implemented via stochastic simulation. Several data analyses are described to illustrate the approach, with resulting neurophysiological insights in some recently generated experimental contexts. Further developments and open questions, both neurophysiological and statistical, are noted.

KEY WORDS: Bayesian computation; Deconvolution; Mixture model; Neural response activity; Parameter identification in mixtures; Quantal synaptic transmission.

1. SCIENTIFIC CONTEXT AND BACKGROUND

The last 15 years or so has seen the development of a major field of study concerning the stochastic characteristics of electrochemical transmission between nerves at synaptic junctions in mammalian and other central nervous systems (Walmsley, Edwards, and Tracy 1987). Fundamental debates in the discipline concern the scope of validity of specific hypotheses about synaptic communication, with variants of a basic, 'quantal' hypothesis being of central interest (Clements 1991; Martin 1966; Walmsley et al. 1988). Much recent experimental and ensuing statistical analysis has been concerned with examining the evidence for or against this simple hypothesis in differing synaptic junctions and under differing environmental conditions (Kullman and Nicoll 1992; Redman 1990; Turner 1987; Turner and Wheal 1991). The relevant statistical issues have been addressed in derived models based on discrete mixture distributions for recorded neural signals, under ranges of partial constraints that impose characteristics either in conformity with quantal hypotheses or allowing specific kinds of departures from such hypotheses (Ling and Tolhurst 1983; Wong and Redman 1980). Our previous statistical work in the area has complemented existing approaches to mixture estimation, such as based on EM computations (Kullman 1989) and entropy formulations (Kullman 1992), by introducing general classes of Bayesian mixture models and their associated analyses (Escobar and West 1995; Turner and West 1993; West and Cao 1993; West and Turner 1994). It is characteristic of these approaches that they adopt a neutral scientific standpoint on the issue of mixture model composition. Like previous, non-Bayesian approaches, the framework has been directed at modeling experimental data using general and flexible classes of mixture models, and then attempting to evaluate the scientific issues of mixture component structure and such issues as the validity of the quantal hypothesis, on a less formal post-hoc basis. This article, in contrast, reports on newer models motivated by the desire to more directly represent the underlying scientific structure, and hence the resulting data configurations. Hierarchical mixture models are structured to reflect ranges of possible and scientifically plausible deviations away from the central and simplifying quantal structure hypothesis, and so provide flexibility in modeling observed nonquantal structure as well as leading to direct and formal assessment of the hypothesis. This allows exploration and reporting of a variety of observed departures from this basic hypothesis and assists in cataloging the diversity of observed excitatory postsynaptic potential (EPSP) phenomena. In developing these models and their analyses, we encounter several issues of wider interest involving in particular prior modeling, Bayesian computation, parameter identifiability, and model comparisons. These issues, among others, are discussed in the context of both theoretical development and several analyses of neurophysiological datasets. Current research frontiers and future prospects are also noted. To begin, we provide some background on the experimental context and examples of current experimental data, leading into the basic modeling framework.

Current neurophysiological experimentation enables the isolation of rather small areas of nerve tissue on large nerve cells in mammalian central nervous systems and in which synaptic connections consist of very small (though uncertain) numbers of transmission sites (Turner and Schlieckert 1990). Electrochemical techniques are used to stimulate nerve tissue to induce neurotransmitter release across identified synapses. Electrical potentials are induced in the cell body of the receiving nerve cell, and estimates of the

Mike West is Professor of Statistics and Director of the Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27708. E-mail: mw@stat.duke.edu; web: http://www.stat.duke.edu. This work was partially supported by National Science Foundation grants DMS-9024793, DMS-9305699, and DMS-9304250. The author acknowledges many useful discussions with Dennis A. Turner and Howard V. Wheal, who provided the data used here (and many other datasets) and with whom collaboration on problems in this area has been the primary stimulus for the reported work. The author also thanks the editor, associate editor, and three referees for their most useful comments on the original version of the article.

Journal of the American Statistical Association
June 1997, Vol. 92, No. 438, Theory and Methods

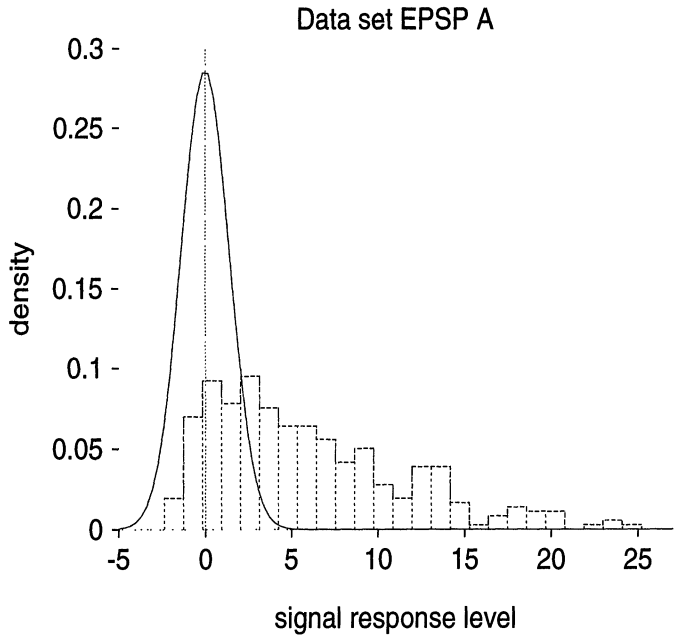

Figure 1. Data From the EPSP-A Experiment. The frame displays a histogram of the EPSP signal data with the estimated normal noise pdf superimposed on the same scale. The superimposed noise density is fitted by matching the mean and variance of the concomitant noise sample.

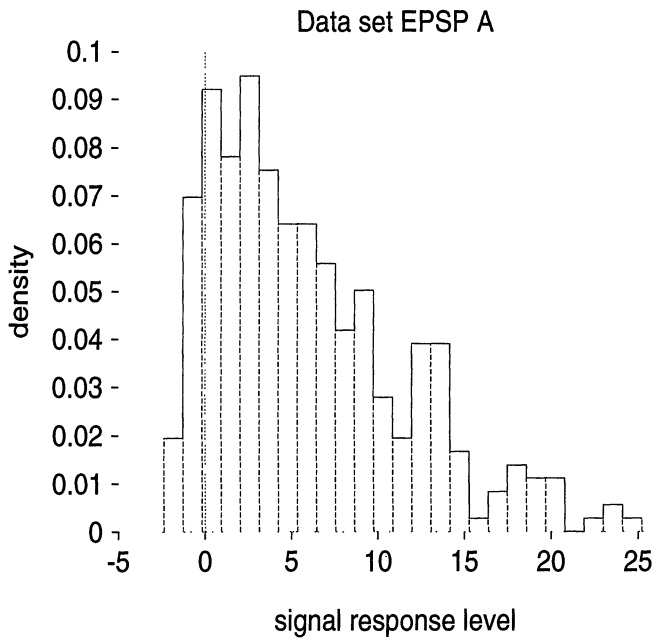

Figure 2. EPSP-A Signal Data, as in Figure 1 but Rescaled for Clarity.

maximum levels of potential reached as a result of transmission are recorded as summary EPSP signals; the units of measurement are millivolts, or meV. This is repeated to produce samples of EPSP levels with sample sizes usually in the hundreds, termed EPSP signal recordings. The signal level achieved on any occasion measures indirectly the level of chemical neurotransmitter released at synaptic sites on the stimulated neuron on that occasion. The measurements are subject to various sources of error: background variability in electrical potentials in the cell, typically including rather minor spontaneous transmission effects, synaptic noise, and noise arising in the physical and electronic recording process. Interspersed between signal recordings are concomitant measurements made under the same experimental protocols but with no active stimulation of the nerve tissue, producing similar samples of measurements on the background noise. These kinds of measurements are simply termed noise recordings; they provide data on noise corrupting whatever structure is evidenced in the EPSP signal data. From the signal and noise datasets together, inferences on the neurophysiological structure underlying neural transmission, as evidenced by the levels of excitation in the (noisy) signal responses, are of interest.

Mixture model structure arises naturally from the science and experimental context, as described in Section 2. To introduce the basic framework, note that neural transmission stems from individual neurotransmitter release sites on the isolated area of the presynaptic neuron; when stimulated, any subset of these sites may release neurotransmitter, and the resulting postsynaptic potential level change induced is the sum of levels corresponding to each of these contributions. Suppose that there are s sites, typically fewer than seven or eight in current experiments. Transmission at any site occurs with some probability, possibly site-specific. Writing $\mu_1, \ldots, \mu_s$ for the release levels for the s sites, it follows that the induced EPSP signal on any occasion reaches a level in a set $\theta_1, \ldots, \theta_k$, say, where each θ_i is a sum of the subset of μ_j values corresponding to sites transmitting. Thus generally there are $k = 2^s$ possible levels, and EPSP signal measurements represent selected values from this set plus contributions from the background experimental and synaptic noise. The mixture structure arises because it is impossible to identify which sites transmit on any occasion. As a result, the signals arise from a mixture

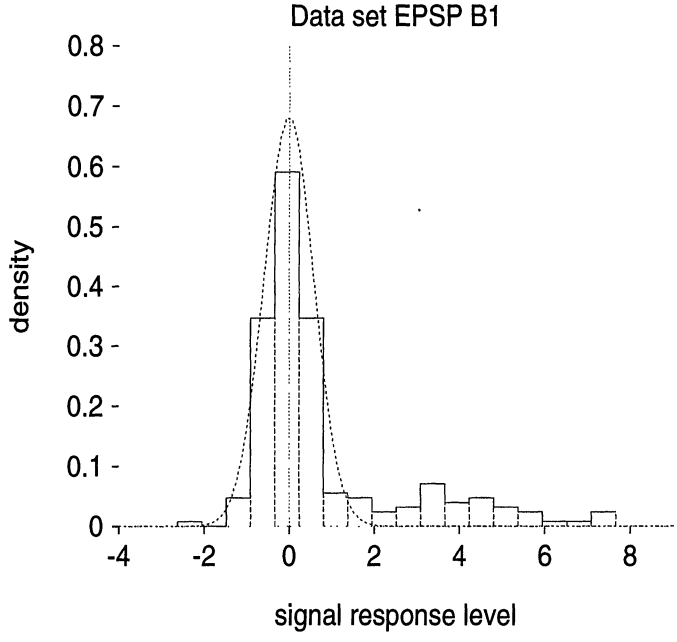

Figure 3. Histogram of Neural Signal Data From the First Pulse of the Paired-Pulse Experiment (EPSP B1) on the Same Synaptic Tissue. There are n = 222 signal observations. The superimposed curve is the estimated normal density of the background noise.

of $k = 2^s$ component distributions. The mixture components are located at the θ_i, themselves linear functions of subsets of the fundamental site release levels μ_j; one key objective is inference about these site levels. The probabilities weighting components are functions of fundamental, site-specific probabilities that govern transmitter release on each occasion; these are also uncertain and to be estimated. The shape of each of the mixture components is that of the background noise distribution. Throughout this work, the noise distribution is assumed to be normal, so the problem is one of inference in a mixture of normals with a common component variance.

This basic model is elaborated on considerably by developing structured hierarchical prior distributions for both the locations and the associated probabilities of the normal components of the mixture. This modeling is based on developments to incorporate qualitative aspects of the underlying science by appropriately describing relationships among the fundamental site-specific release levels and release probabilities. The number s of underlying release sites is also uncertain, which induces uncertainty about the number k of mixture components; our work addresses this issue through prior modeling and sensitivity analyses. For now, the key point is that the mixture model arises from the scientific context; nonnormal features evidenced in signal data configurations thus are attributable to the mixture parameters. In my models, I negate the original negative excitatory potential changes so that the μ_j are nonnegative. Hence EPSP signal data configurations are expected to appear as positively skewed, possibly with multiple modes located near positive component means. The examples displayed in Figures 1–4 are typical of current experiments.

I follow most previous works in assuming normally distributed noise, for two main reasons. First, a primary focus here is to present novel and appropriate models and analyses for the signal recordings, extending previous approaches based on normal noise distributions. Considering nonnormal noise distributions would unduly complicate and cloud the key issues here—the objective of closer and more appropriate modeling of the underlying scientific structure of the signal data. Second, ranges of existing EPSP datasets and others being generated by neurosurgeons have noise recordings that closely conform to normality assumptions. For the future, it should be noted that some EPSP experiments are quite evidently subject to nonnormal errors and others more subtlety so; hence appropriately capturing the noise characteristics is clearly important in some EPSP signal analyses. I am currently involved in developing extensions of the models here to flexibly incorporate nonnormal noise distributions, extending previous work in signal models (Cao and West 1996); such extensions will be reported elsewhere.

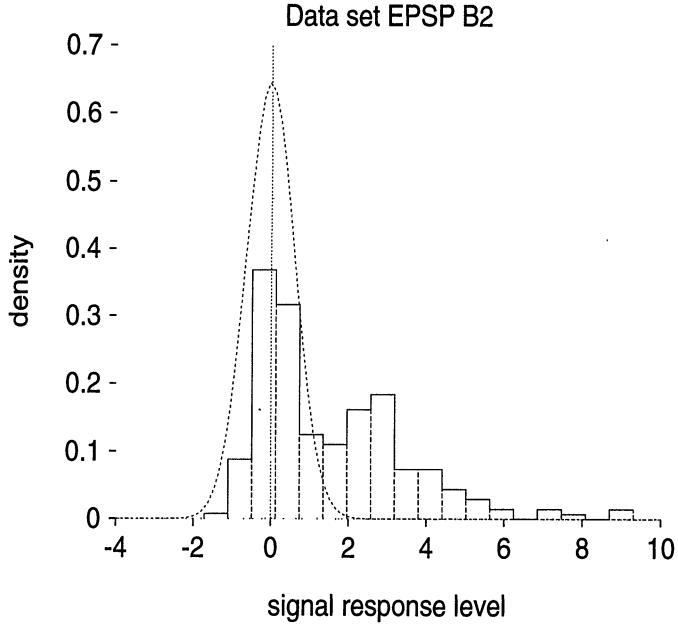

Figure 4. Histogram of Neural Signal Data From the Second Pulse of the Paired-Pulse Experiment (EPSP B2) on the Same Synaptic Tissue. As in Figure 3, there are n = 222 signal observations and the estimated normal density of the background noise is superimposed.

A first example EPSP dataset appears in Figures 1 and 2. The normal pdf determined by the sample moments of the corresponding noise measurements is superimposed on the signal histogram in Figure 1; the signal dataset alone appears in Figure 2. The range and shape of the signal histogram relative to the noise indicate an overall mixture of quite a few components with appreciable probabilities, consistent with more than one active release site. This is a typical data configuration and a typical sample size of $n = 325$. This dataset comes from an experiment with synaptic tissue from a region of the hippocampus of rat (Turner, Isaac, Chen, Stockley, and Wheal 1995b). The datasets to follow come from similar regions.

Two additional example EPSP datasets appear in Figures 3 and 4. In each case the normal pdf determined by the sample moments of the corresponding noise measurements is superimposed on the signal histogram. These data were obtained in a paired-pulse experiment performed by Howard Wheal of Southampton University. A paired-pulse experiment is designed to investigate changes or adaptations in the neural response characteristics at a single synaptic junction due to previous potentiation (e.g., Kullman and Nicoll 1992). In this case the tissue is stimulated as usual to generate the first EPSP dataset, both noise and signal recordings; this is EPSP-B1 in Figure 3. Then, following potentiation, the experiment is rapidly repeated, generating EPSP-B2 of Figure 4. Inference about differences in release levels, and in the stochastic nature of neurotransmitter release more generally, are relevant to physiological theories of adaptation and evolution of nervous systems and are of particular interest in connection with growth (or deterioration) in function in mammalian nervous systems and brains (Kullman and Nicoll 1992). So it is of interest here to compare inferences made for the two cases, asking questions about how the stochastic mechanisms governing transmission may have changed between the two cases as a result of recent potentiation. The datasets displayed are atypical in the small sample sizes, which will evidently be a limiting factor in inferences about site parameters, but are typical of recent paired-pulse experiments of Wheal, which are designed to

stimulate rather few neural sites with low release probabilities.

Section 2 describes the basic modeling framework introduced in this article and interesting special cases. This includes detailed and quite intricate structure for prior distributions for parameters representing individual transmission site characteristics. Section 3 describes Bayesian analyses via stochastic simulation, discussing various computational issues and critical parameter identification questions; the Appendix contains summary technical details for the simulation analysis. Section 4 summarizes some analyses of the EPSP-A data, exploring modeling and inference issues of general concern and discussing the specific neural dataset. Section 5 discusses analyses of the paired-pulse datasets with more of a focus on the scientific issues. Section 6 explores a scientifically very important and recent model extension that introduces significant complications into the analysis and raises some issues of analysis in the context of the EPSP-B2 data. Section 7 concludes with some summary comments and indications of current and expected modeling developments.

2. MODELS FOR SITE-SPECIFIC RELEASE CHARACTERISTICS

2.1 Neural Signal and Noise Models

Replicate measures of maximum EPSP signal levels are assumed to be independently generated according to an underlying stochastic, synaptic site-specific mechanism described as follows. At the identified synaptic connection, the experiment isolates a very small region on the axon branch of a presynaptic neuron, and stimulus of the neuron leads to chemical neurotransmitter release at some or all of a few minute synaptic transmission sites. The sum across sites of the neurotransmitter released on any occasion induces the potential change in the receiving neuron. This is measured, subject to noise. Each of the $s > 0$ release sites transmits to a site-specific maximum level independently of all other sites. There is no identification of individual sites (e.g., such as might arise were one able to physically mark sites in the synaptic tissue and identify some individual transmissions) so the sites are arbitrarily labeled $1, \ldots, s$. Site-by-site, it is assumed that on any trial:

- sites "fire" independently, with individual, site-specific chances of firing on any occasion
- a transmitting site produces a site-specific, fixed packet or "quantum" of neurotransmitter
- recorded maximum EPSP levels present the sums of potentials induced by the sites transmitting, with additive synaptic and experimental noise.

In any experiment individual transmitter release probabilities and transmission levels are assumed to be fixed. Changes in these characteristics over time due to forced experimental and environmental changes is of later interest, and evaluating such changes is one of the guiding motivations in developing these models; see Section 5 for illustration. It is stressed that these assumptions incorporate and relax standard, consensus assumptions about neural function underlying EPSP activity. Symbolically, recorded EPSP levels are represented by $\mathbf{y} = (y_1, \ldots, y_n)'$, where y_i is the level on trial i; the sample size n is typically a few hundreds. Given s sites assumed, site j transmits on any trial with probability π_j, independently of other sites and independently across trials. The level of transmission on firing is the site-specific quantity μ_j. Physically, EPSP levels are measured via intracellular probes as electrical potentials induced on cell bodies, and responses are in negative potentials in units of meV. By convention, the readings are negated so that the μ_j must be nonnegative values. The recordings are subject to background synaptic and experimental noise, including both electronic distortions in recording the cellular potential and errors induced in computing the estimated maximum potential levels. These noise sources combine to produce additive errors, assumed to be approximately normally distributed independently across trials, as discussed in Section 1; write v for the variance of these normal errors. Concomitant noise measurements are available to assess this and to provide prior information relevant to v; differences in cell body potentials measured without experimental stimulus provide these measurements. It is stressed that the noise affecting signal recordings has the same distribution, of variance v, as the raw noise recordings, there being no scientific or experimental reason to assume otherwise. A systematic bias in the signal recordings is also modeled; although the associated noise measures are typically close to zero mean, the signal recordings are often subject to a small but nonnegligible potential shift induced by the experimental procedures. This shift is called m. Then the following basic signal model is obtained.

For $i = 1, \ldots, n$, the signal observations are conditionally independent,

$$y_i \sim N(y_i|\theta_i, v) \quad \text{with} \quad \theta_i = m + \sum_{j=1}^{s} z_{ij}\mu_j,$$

and where the z_{ij} are latent unobserved site transmission indicators,

$$z_{ij} = \begin{cases} 1, & \text{if site } j \text{ transmits on trial } i \\ 0, & \text{otherwise.} \end{cases}$$

Under the model assumptions, these indicators are conditionally independent Bernoulli quantities with $P(z_{ij} = 1) = \pi_j$ for each site j and all trials i.

Thus EPSP data are drawn from a discrete mixture of (at most) $k = 2^s$ normal components of common variance, induced by averaging normals determined by all possible combinations of the z_{ij} and weighted by the corresponding chances of those combinations. On any trial, the normal component selected is determined by the column s vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{is})'$ realized on that trial by individual sites firing or not firing; the selected component has mean $\theta_i = m + \mathbf{z}_i'\boldsymbol{\mu}$, where $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_s)$, and is selected with chance $\prod_{j=1}^{s} \pi_j^{z_{ij}}(1 - \pi_j)^{1-z_{ij}}$. Specifically, and conditional on all model assumptions and parameters, the y_i are conditionally independently drawn from the distribu-

tion with density

$$p(y_i|\boldsymbol{\mu},\boldsymbol{\pi},m,v) = \sum_{\mathbf{z}_i} p(\mathbf{z}_i|\boldsymbol{\pi})p(y_i|\boldsymbol{\mu},\mathbf{z}_i,m,v)$$

$$= \sum_{\mathbf{z}_i} \left\{ \prod_{j=1}^{s} \pi_j^{z_{ij}}(1-\pi_j)^{1-z_{ij}} \right\} \times N(y_i|m+\mathbf{z}_i'\boldsymbol{\mu},v), \quad (1)$$

where the s vector $\mathbf{z}_i$ ranges over all 2^s possible values. If there are common values among the site levels μ_j, then the mixture distribution function will be equivalent to one with fewer than 2^s components. Also, if either $\mu_j = 0$ or $\pi_j = 0$, then site j disappears from the model and the distribution reduces to a mixture of at most 2^{s-1} components. Additional cases of special interest include:

a. So-called compound binomial models, in which the transmission levels are equal, all at a basic quantum level μ_0, but release probabilities differ. The aggregate transmission levels then run between m and $m + s\mu_0$ and the mixture of 2^s normal components effectively reduces to a mixture of just $s + 1$, with associated weights determined by the discrete compound binomial resulting from the distinct chances $\pi_1, \ldots, \pi_s$ across sites. These kinds of special cases have received considerable attention in the literature (Kullman 1992; Walmsley et al. 1988).
b. Precise quantal-binomial models, in which the site levels are equal and the release probabilities are constant too, $\pi_j = \pi_0$. Now the mixture reduces to one with distinct normal means $m + j\mu_0$ $(j = 0, \ldots, s)$, and binomial weights $\binom{s}{j}\pi_0^j(1-\pi_0)^{s-j}$ (Clements 1991; Martin 1966; Redman 1990).

In previous work, following others, I modeled EPSP data in the framework of standard mixture models for density estimation from a particular Bayesian viewpoint (see, e.g., West and Cao 1993 and West and Turner 1994, and Escobar and West 1995 for statistical background). In these models the component means and weights are essentially unrestricted, rather than being modeled directly as functions of the underlying synaptic parameters $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$. A significant drawback is that it is then in general very difficult to translate posterior inferences about unrestricted normal mixture model parameters to the underlying $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$ parameters of scientific interest, especially in the context of uncertainty about s. One simple example of West and Turner (1994) shows how this can be done; that is a rare example in which the data appear to be consistent with the full quantal hypothesis, so that the convoluted process of backtracking from the mixture model to the underlying site-specific parameters is accessible. However, I have encountered very few datasets in which this is the case, and this inversion process is generally difficult. Hence the direct approach, inferring the neural parameters directly, is developed.

2.2 Prior Distributions for Synaptic Parameters

Model completion requires specification of classes of prior distributions for the determining parameters $\boldsymbol{\mu}$, $\boldsymbol{\pi}$, m, and v for any given s. Assessment of reasonable values of s, as well as values of the $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$ quantities for any given s, is part of the statistical inference problem. From a technical viewpoint, uncertainty about s can be formally included in the prior distribution, so as to provide posterior assessment of plausible values. The models below allow this. From the viewpoint of scientific interpretation, however, inference about the site parameters are best made conditional on posited values of s, and then issues of sensitivity to the number of sites explored. Development and examples explore these issues in detail.

Begin by conditioning on a supposed number of sites s. Current implementation assumes priors for all model parameters which, though rather intricately structured in certain dimensions to represent key qualitative features of the scientific context, are nevertheless inherently uniform in appropriate ways, providing reference initial distributions. Note that other classes of priors may be used (some obvious variations are mentioned later); however, those used here are believed to be appropriately vague or uninformative so that resulting posterior distributions provide benchmark or reference inferences that may be directly compared to other approaches (see, e.g., Kullman 1989, 1992).

In addressing uncertainty about s alone, I make the observation that models with fewer than s sites are implicitly nested within a model having s sites. To see this, constrain some $r > 0$ (arbitrarily labeled) site levels μ_j to be 0; then (1) reduces to precisely the same form based on $s - r$ sites, whatever the value of the π_j corresponding to the zeroed μ_j. Zeroing some μ_j simply confounds those sites with the noise component in the mixture—a site transmitting a zero level with any probability simply cannot be identified. Hence assessing whether or not one or more of the site levels are 0 provides assessment of whether or not the data support fewer sites than assumed. This provides a natural approach to inference on the number of active sites, assuming that the model value s is chosen as an upper bound. Note that a similar conclusion arises by considering $\pi_j = 0$ for some indices j; that is, an inactive site may have a zero release probability rather than (or as well as) a zero release level. However, the zeroing of a site probability induces a degeneracy in structure of the model and obviates its use as a technical device for inducing a nesting of models with fewer than s sites in the overall model. Hence my models define inactive sites through zeros among the μ_j, restricting the π_j to nonzero values, however small. Thus $\mu_j = 0$ for one or more sites j is the only way that the number of active sites may be smaller than the specific s.

2.2.1 General Structure. For a specified s, analyses reported here are based on priors with the following structure. Quantities $\boldsymbol{\mu}$, $\boldsymbol{\pi}$, (m, v) are mutually independent. I describe classes of marginal priors for $\boldsymbol{\mu}$ and $\boldsymbol{\pi}$ separately; each involves certain hyperparameters that themselves are subject to uncertainty described through hyperpriors, in a typical hierarchical modeling framework, and these hyper-

parameters also are assumed to be mutually independent. To anticipate later development hyperparameters denoted by q and a are associated with the prior for $\boldsymbol{\mu}$, a single quantity b determines the prior for $\boldsymbol{\pi}$, and the joint prior is then of the form

$$\begin{aligned} &p(\boldsymbol{\mu}, \boldsymbol{\pi}, m, v, q, a, b) \\ &\quad = p(\boldsymbol{\mu}, \boldsymbol{\pi}, m, v|q, a, b)p(q, a, b) \\ &\quad = p(\boldsymbol{\mu}|q, a)p(\boldsymbol{\pi}|b)p(m, v)p(q)p(a)p(b). \end{aligned} \tag{2}$$

The component densities here are now described in detail. Two comments on notation: First, conditioning statements in density functions include only those quantities that are required to determine the density, implicitly indicating conditional independence of omitted quantities; second, for any vector of h quantities $\mathbf{x} = (x_1, \ldots, x_h)'$ for any $j \leq h$, the notation $\mathbf{x}_{-j}$ represents the vector $\mathbf{x}$ with x_j removed; that is, $\mathbf{x}_{-j} = \mathbf{x} - \{x_j\} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_h)'$.

2.2.2 Priors for $\boldsymbol{\mu}$ and Hyperparameters q, a. I develop a general classes of priors for $\boldsymbol{\mu}$ and comment on various special cases. The class has the following features:

a. a component baseline uniform distribution for each μ_j over a prespecified range $(0, u)$, with a specified upper bound u
b. components introducing positive prior probabilities at 0 for each of the μ_j to permit assessment of hypotheses that fewer than the chosen (upper bound) s are actually nonzero, and hence to infer values of the number of active sites
c. components permitting exact common values among the elements of $\boldsymbol{\mu}$ to allow for the various special cases of quantal transmission, and specifically the questions of whether or not pairs or subsets of sites share essentially the same quantal transmission level.

These are developed as follows. Let $F(\cdot)$ be distribution on $(0, u)$, having density $f(\cdot)$, for some specified upper bound u. Write $\delta_0(x)$ for the Dirac delta function at $x = 0$ and $U(\cdot|a, b)$ for the continuous uniform density over (a, b). Then suppose that the μ_j are *conditionally independently* drawn from the model

$$(\mu_j|F, q) \sim q\delta_0(\mu_j) + (1 - q)f(\mu_j)$$

for some probability q. Write h for the number of nonzero values among the μ_j, so that h is the number of active sites. Then h is binomial, $(h|s, q) \sim \text{Bn}(h|s, 1 - q)$ with mean $s(1 - q)$, independent of F. If F were uniform, say $U(\cdot|0, u)$, then this prior would neatly embody the first two desirable features a and b just described. Note the two distinct cases:

- Setting $q = 0$ implies that $h = s$ is the assumed number of active sites, so then inference proceeds conditional on $\mu_j > 0$ for each j.
- Otherwise, restricting to $q > 0$ allows for assessment of the number of active sites, subject to the specified upper bound s. In practice I assign a hyperprior to q in the case $q > 0$. The class of beta distributions is conditionally conjugate, and the uniform prior suggests itself as a reference, $q \sim U(q|0, 1)$. One immediate, and nice, consequence of a uniform prior is that the resulting prior for the number of nonzero values among the μ_j has (averaging the binomial with respect to q) a discrete uniform prior over $0, 1, \ldots, s$. This is a suitably vague and unbiased initial viewpoint with respect to the number of active sites. Other beta priors may be explored, of course. One specific choice used in the current work is $(q|s) \sim \text{Be}(q|s - 1, 1)$; note the explicit recognition of dependence on the specified values of s. The reasoning behind this choice is as follows. First, the current EPSP experiments are designed to isolate rather small numbers of sites, down to just a few (say one–four) from the viewpoint of scientific intent and expectation. So s values up to 7 or 8 may be explored, but lower values are typical. Whatever value of s is chosen, h is expected to be in the low integers, thus guiding the choice of the prior for q to induce a prior for h favoring smaller values. Consider values of s in the relevant range $3 \leq s \leq 8$, or so. Then, integrating $p(h|s, q)$ with respect to the specific prior $(q|s) \sim \text{Be}(q|s - 1, 1)$ yields a distribution $p(h|s)$ that is almost completely insensitive to s, having a diffuse and decreasing form as h increases, and with $E(h|s) = 1$ for any such s. This specific beta prior for q thus has the attractive feature of consistency with a scientifically plausible prior $p(h|s) \approx p(h)$, incorporating the scientific view of likely small numbers of active sites and almost independently of the upper bound s specified.

This structure provides a baseline uniform prior for release levels, together with the option for allowing a smaller number of sites than the s specified. So far, however, there is no explicit recognition of the special status in the scientific area of quantal hypotheses as represented through common values among the μ_j. If $F(\cdot)$ is a continuous distribution, then the prior implies that the nonzero μ_j are distinct. Although this might allow arbitrarily close μ_j values, it is desirable to have the opportunity to directly assess questions about common values, and perhaps subgroups of common values, in terms of posterior probabilities. There is also a significant technical reason (discussed in Section 3 in connection with issues of parameter identification) that calls for a prior that gives positive probability to exact equality of collections of the nonzero μ_j. Thus I extend the prior structure so-far discussed to provide this. I do this using a standard Dirichlet model, fully described in Appendix A. Specifically, a Dirichlet process prior for F induces a discrete structure that gives positive prior probability to essentially arbitrary groupings of the set of nonzero μ_j into subsets of common values; this is a general framework that permits varying degrees of partial quantal structure, from the one extreme of completely distinct values to the other of one common value. This structure is most easily appreciated through the resulting set of complete conditional posterior distributions for each of the μ_j given $\boldsymbol{\mu}_{-j}$. As detailed in

Appendix A, these are defined by

$$p(\mu_j|\boldsymbol{\mu}_{-j}, q, a) = q\delta_0(\mu_j) + (1-q) \times \left\{ r_j U(\mu_j|0,u) + (1-r_j)h_j^{-1} \sum_{i \in N_j} \delta_{\mu_i}(\mu_j) \right\}, \quad (3)$$

where h_j is the number of nonzero elements of $\boldsymbol{\mu}_{-j}$, N_j is the corresponding set of indices, $N_j = \{i|\mu_i > 0, i = 1,\ldots,s; i \neq j\}$, and $r_j = a/(a+h_j)$. The hyperparameter a is subject to uncertainty and is included in the analysis using existing approaches for inference on precision parameters in Dirichlet models, developed by West (1992) and illustrated by Escobar and West (1995). As these authors showed gamma priors for a, or mixtures of gamma priors, are natural choices. The current application uses diffuse gammas models.

In summary, (3) shows explicitly how site level μ_j may be 0 (implying an inactive site), may take a new nonzero value, or may be equal to one of the nonzero values of other sites. The roles of hyperparameters q and a are evident in this equation. With this structure, prior components $p(\boldsymbol{\mu}|q,a)p(q)p(a)$ of the full joint prior in Equation (2) have been defined.

2.2.3 Priors for π and Hyperparameter b. The structure of the prior for $\boldsymbol{\pi}$ parallels, in part, that of $\boldsymbol{\mu}$ in allowing common values. The detailed development for $\boldsymbol{\mu}$ can be followed through with the same reasoning about a baseline prior and a structure to induce positive probabilities over subsets of common values. I restrict the discussion to positive release probabilities, as discussed earlier, so that the prior structure for the π_j is simpler in this respect. Specifically, assume that the π_j are independently drawn from a distribution on (0, 1) assigned a Dirichlet process prior. Take Dirichlet precision $b > 0$ and base measure $bU(\cdot|0,1)$. Then, as in the development for $\boldsymbol{\mu}$, the full joint prior $p(\boldsymbol{\pi}|b)$ is defined by its conditionals

$$(\pi_j|\boldsymbol{\pi}_{-j}, b) \sim wU(\pi_j|0,1) + (1-w)(s-1)^{-1} \sum_{i=1, i\neq j}^{s} \delta_{\pi_i}(\pi_j) \quad (4)$$

for each $j = 1,\ldots,s$, and where $w = b/(b+s-1)$. As for the site levels, the induced posteriors will now allow inference on which sites may have common release probabilities, as well as on the precise values of such probabilities.

Nonuniform beta distributions might replace the uniform baseline in application, and they are particularly relevant in analyses of more recent datasets generated with a view toward engendering rather low levels of excitation consistent with expected low levels of release probabilities. The analyses reported herein retain the uniform prior, for illustration and so as to avoid question of overtly biasing toward lower or higher values. Note also that the possibility of nonzero release probabilities is explicitly excluded, although some or all may be very small, and so $\pi_j = 0$ is ruled out as a device for reducing the number of active sites. The number of active sites is completely determined by zero values among the μ_j, as discussed earlier.

As with a in the model for the site levels, the hyperparameter b is assigned a prior, and again a diffuse gamma prior is natural. With this structure, the additional prior components $p(\boldsymbol{\pi}|b)p(b)$ of the full joint prior in Equation (2) have been defined.

2.2.4 Priors for Noise Moments m and v. In most experiments one can anticipate a possible systematic bias m in both noise and signal recordings, induced by the direct measurement of cell membrane potential. This is typically very small relative to induced EPSP levels, and the raw noise recordings provide data on which to assess this. In addition, the noise measurements inform on background variability v; that is, they are drawn from the $N(\cdot|m,v)$ distribution. In my analyses, the prior for m and v as input to the signal analysis is simply the posterior from a reference analysis of the noise data alone; that is, a standard conjugate normal, inverse gamma distribution based on the noise sample mean, variance, and sample size.

With this structure, the final component $p(v,m)$ of the full joint prior in Equation (2) has been defined.

3. POSTERIOR DISTRIBUTION AND COMPUTATION

Calculation of posterior distributions is feasible, as might be expected, via variants of Gibbs sampling (see, e.g., Gelfand and Smith 1990, and Smith and Roberts 1993). The specific collection and sequence of conditional posterior distributions used to develop simulation algorithms are described in some mathematical detail in Appendix B and briefly summarized here.

A key step is to augment the data with the latent site transmission indicators $\mathbf{z}$; thus the sampling model is expanded as

$$p(\mathbf{y}, \mathbf{z}|\boldsymbol{\mu}, \boldsymbol{\pi}, m, v, q, a, b) = p(\mathbf{y}|\boldsymbol{\mu}, \mathbf{z}, m, v)p(\mathbf{z}|\boldsymbol{\pi}) = \left\{ \prod_{i=1}^{n} N(y_i|m + \mathbf{z}_i'\boldsymbol{\mu}, v) \right\} \times \left\{ \prod_{j=1}^{s} \prod_{i=1}^{n} \pi_j^{z_{ij}} (1-\pi_j)^{1-z_{ij}} \right\},$$

providing various conditional likelihood functions that are neatly factorized into simple components. The full joint posterior density for all model parameters together with the uncertain indicators $\mathbf{z}$ has the product form

$$p(\boldsymbol{\mu}, \boldsymbol{\pi}, \mathbf{z}, m, v, q, a, b|\mathbf{y}) \propto p(\boldsymbol{\mu}|q,a)p(q)p(a)p(\mathbf{z}|\boldsymbol{\pi}) \times p(\boldsymbol{\pi}|b)p(b)p(m,v)p(\mathbf{y}|\boldsymbol{\mu}, \mathbf{z}, m, v), \quad (5)$$

where the component conditional prior terms are as described in the previous section.

3.1 Conditional Posteriors and Markov Chain Monte Carlo

The posterior (5) yields a tractable set of complete conditional distributions characterizing the joint posterior, and hence leads to implementation of Gibbs sampling based on this structure. Each iteration of this Markov Chain Monte Carlo (MCMC) sampling scheme draws a new set of all parameters and latent variables by sequencing through the conditionals now noted; Appendix B gives full details and derivations.

1. Sampling site levels $\boldsymbol{\mu}$ proceeds by sequencing through $j = 1, \ldots, s$, at each step generating a new value of μ_j given the latest sampled values of $\boldsymbol{\mu}_{-j}$ and all other conditioning quantities. For each j, this involves sampling a posterior that is a simple mixture of several point masses with a truncated normal distribution. Sampling efficiency is improved using variations of so-called configuration sampling for the discrete components, as introduced by MacEachern (1994) and developed by West, Müller, and Escobar (1994).

2. Sampling site release probabilities $\boldsymbol{\pi}$ similarly proceeds by sequencing through $j = 1, \ldots, s$, at each step generating a new value of π_j given the latest sampled values of $\boldsymbol{\pi}_{-j}$ and all other conditioning quantities. For each j, this involves sampling a mixture of discrete components with a beta component. Again, configuration sampling improves simulation efficiency.

3. Sampling transmission indicators $\mathbf{z}$ involves a set of n independent draws from conditional multinomial posteriors for the n individual binary 2^s-vectors $\mathbf{z}_i$, $i = 1, \ldots, n$. These simulations are easily performed.

4. Sampling the systematic bias quantity m involves a simple normal draw.

5. Sampling the noise variance v involves sampling an inverse gamma posterior.

6. Sampling the hyperparameter q is also trivial, simply involving a draw from an appropriate beta distribution.

7. Sampling the hyperparameters a and b follows West (1992) and Escobar and West (1995) and involves a minor augmentation of the parameter space, with simple beta and gamma variate generations.

Iterating this process produces sequences of simulated values of the full set of quantities $\phi \stackrel{\text{def}}{=} \{\boldsymbol{\mu}, \boldsymbol{\pi}, \mathbf{z}, m, v, q, a, b\}$, which represent realizations of a Markov chain in ϕ space whose stationary distribution is the joint posterior (5) characterized by the conditionals. The precise sequence 1–7 of the foregoing conditional posteriors (detailed more fully in Appendix B) is the sequence currently used in simulation; at each step, new values of the quantities in question are sampled based on current values of all quantities required to determine the conditional in question. Convergence is assured by appeal to rather general results, such as those of Tierney (1994, especially thm. 1 and cor. 2) relevant to the models here; this ensures that successive realizations of $\phi = \{\boldsymbol{\mu}, \boldsymbol{\pi}, \mathbf{z}, v, q, a, b\}$ generated by this Gibbs sampling setup eventually resemble samples from the exact posterior in (5). Irreducibility of the resulting chain is a consequence of the fact that all full conditionals defining the chain are everywhere positive. Practical issues of convergence are addressed by experiments with several short runs from various starting values. Important practical issues of convergence is discussed in the next section and in the EPSP data analyses that follow.

3.2 Parameter Identification and Relabelling Transmission Sites

The model structure as described is not identifiable, in the traditional sense of parameter identification especially associated with mixture models (Titterington, Smith, and Makov 1985, chap. 3). This is obvious in view of the lack of any physical identification of the neural transmission sites arbitrarily labelled $j = 1, \ldots, s$; for example, π_1 is the release probability for the first site in the labelling order, but this could be any of the actual release sites. As a result, the model is invariant under arbitrary permutations of the labels on sites, and likelihood functions are similarly invariant under permutations of the indices of the sets of parameters (μ_j, π_j). Also, as the priors for the μ_j and π_j separately are exchangeable, corresponding posteriors are also invariant under labelling permutations.

To impose identification requires a physically meaningful restriction on the parameters. Perhaps the simplest, and most obvious, constraint is to order either site levels or site release chances. This can be done with the site release levels. In posterior inferences, relabel so that site j corresponds to release level μ_j, the jth largest value in $\boldsymbol{\mu}$. Because the priors, and hence the posteriors, include positive probabilities on common values, ties are broken by imposing a subsidiary ordering on the release probabilities for (ordered) sites with common release levels. Thus if μ_3, μ_4, and μ_5 happened to be equal, the sites relabelled 3, 4, and 5 would be so chosen in order of increasing values of their chances π_j; were these equal too, as permitted under the priors and hence posteriors here, then the labelling is arbitrary and there are (in this example) three indistinguishable sites: common release levels and common release probabilities. The posterior analysis required to compute corresponding posterior inferences is extremely simple in the context of posterior simulations; each iteration of the Gibbs sampling analysis produces a draw of $(\boldsymbol{\mu}, \boldsymbol{\pi})$ (among other things). Given a draw, a separate vector $\boldsymbol{\mu}^*$ is created containing the *ordered* values of $\boldsymbol{\mu}$, then a second vector $\boldsymbol{\pi}^*$ is created containing the elements of $\boldsymbol{\pi}$ rearranged to correspond to the ordering of $\boldsymbol{\mu}$ to $\boldsymbol{\mu}^*$. If subsets of contiguous values in $\boldsymbol{\mu}^*$ are common, then the corresponding elements in $\boldsymbol{\pi}^*$ are rearranged in increasing order themselves. Through iterations, repeat draws of $(\boldsymbol{\mu}^*, \boldsymbol{\pi}^*)$ are saved, building up samples from the required posterior distribution that incorporates the identification of sites.

Note that alternatively, the model could be identified by a primary ordering on the π_j, rather than on the μ_j. In data analyses I routinely explore posterior summaries identified each (i.e., ordered by μ_j and ordered by π_j), as each is only a partial summary of the full posterior. This is relevant

particularly in cases when it appears that the data are in conformity, at least partially, with quantal structure. Some discussion follows in data analyses in later sections.

In connection with this ordering issue, the quantal structure of the priors for μ and π is very relevant, as alluded to earlier. Consider an application in which the data support common, or very close, values among the μ_j. Suppose that the prior is simply uniform, with no chance that consecutive values of the ordered μ_j are equal. Then imposing the ordering for identification results in a distortion of inferences—values that should be judged close, or equal, are "pushed apart" by the imposed ordering. This undesirable effect is ameliorated by the priors used, allowing exact equality of neighboring levels. The same applies to the π_j.

Note that the original, unidentified parameterization for model specification is retained, and the simulation analysis operates on the unconstrained posteriors. This has a theoretical benefit in allowing easy and direct application of standard convergence results (Tierney 1994). More practically, it has been my experience that imposing an identifying ordering on parameters directly through the prior distribution hinders convergence of the resulting MCMC on the constrained parameter space. This arises naturally as the ordering induces a highly structured dependence between parameters, relative to the unconstrained (albeit unidentified) parameter space. In addition, one rather intriguing practical payoff from operating the MCMC on the unrestricted space is an aid in assessing convergence of the simulation. This is due to the exact symmetries in posterior distributions resulting from permutation invariance, partly evidenced through the fact that marginal posteriors $p(\mu_j, \pi_j|y)$ are the same for all j. Hence summaries of marginal posteriors for, say, the μ_j should be indicative of the same margins for all j; the same holds for the π_j. However, in many applications convergence will be slow. One reason for this is that in some models and with some observed signal datasets, the posterior will concentrate heavily in widely separate regions of the parameter space, possibly being multimodal with well-separated modes. Such posteriors are notorious in the MCMC literature, and the Gibbs sampling routine may get "stuck" in iterations in just one (or typically several) region of high posterior probability, or around a subset of modes. Some of the examples that follow illustrate this phenomenon, and there is certainly need for further algorithmic research to provide faster alternatives to the raw Gibbs sampling developed here (possibly based on some of the methods mentioned in Besag and Green 1993 or Smith and Roberts 1993, for example). However, this is not a problem as far as the ordered site levels, and the correspondingly reordered release probabilities, are concerned, due to the symmetries in the unconstrained posterior. This is because posterior samples from one region of the (μ, π) space may be reflected to other regions by permuting sites indices, while the resulting ordered values remain unchanged. Hence the unconstrained sampling algorithm may converge slowly as far as sampling the posterior for (μ, π), but the derived sequence for (μ^*, π^*) converges much faster. This is discussed further in data analysis in Section 4.

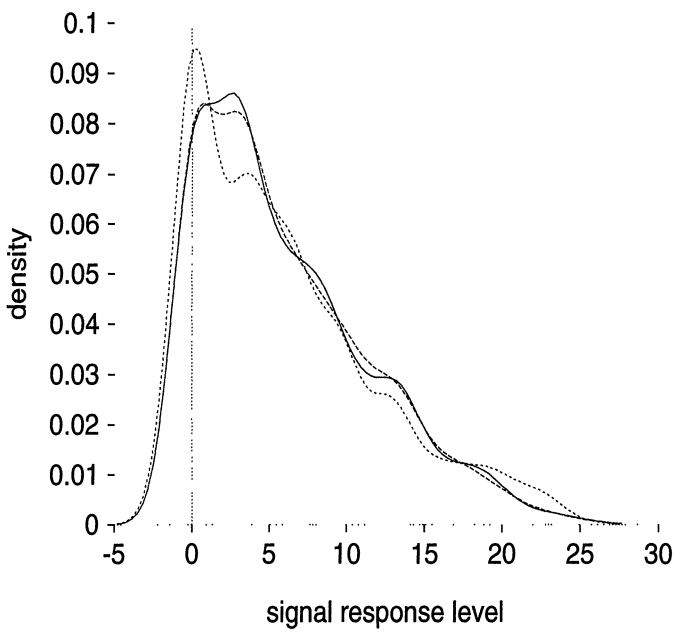

Figure 5. Summaries From Analyses of EPSP-A Data: Predictive Densities From Analyses in Models With Three Sites (Dotted Line), Four Sites (Dashed Line), and Five Sites (Full Line).

Due to the lack of identification, actual output streams from MCMC simulations exhibit switching effects, as parameter draws jump between the modes representing the identification issue. For example, in a study supporting two distinct site release levels with values near 1 and 2 meV, the simulated series of the unidentified $\mu_1(\mu_2)$ tends to vary around 1 meV (2 meV) for a while, then randomly switch to near 2 meV (1 meV) for a while, and so on. My experience is that across several analyses of different EPSP datasets, convergence is generally "clean" in the sense that, following what is usually a rapid initial burn-in period, the output streams remain stable and apparently stationary between points of switching between these posterior modes. Informal diagnostics based on a small number of short repeat runs from different starting values verify summarized analyses from one final, longer run. In addition, the theoretical convergence diagnostic based on the symmetry of the unidentified posterior is used in final analysis, as illustrated later.

4. ANALYSES OF THE EPSP-A DATASET

Some summaries of initial analyses of the EPSP-A are displayed in Figures 5–10. First, three separate analyses were performed using models with $s = 3$, 4, and 5, respectively; to do this, $q = 0$ was set in the prior $p(\mu|q, a)$ to fix the number of sites at the specified s value in each case. The prior for the μ_j has $u = 20.0$. Each MCMC simulation was run for 45,000 iterations; summaries of the last 40,000 are reported. Informal assessment of convergence was performed as discussed at the end of Section 3.

Figure 5 displays posterior predictive density functions from the three analyses—Bayesian density estimates, of use in graphical assessment of model fit. These are simply computed as Monte Carlo averages of the density func-

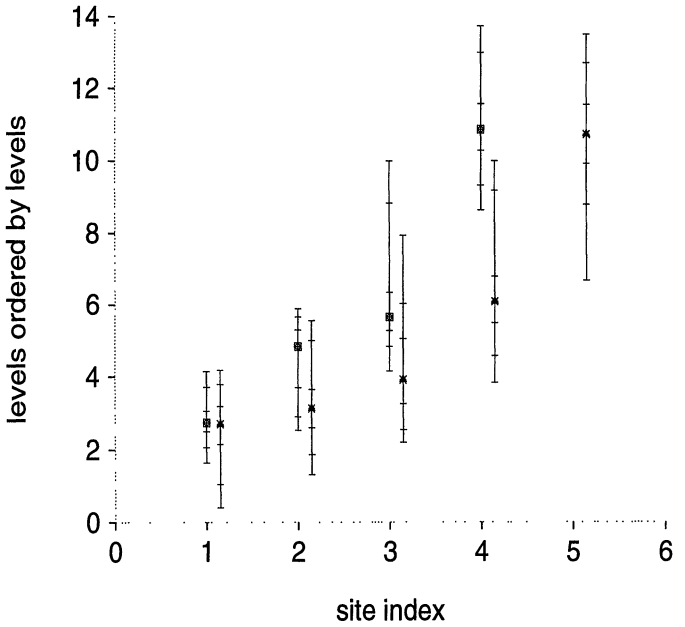

Figure 6. Summaries From Analyses of EPSP-A Data: At Each Site Index j, Posterior Intervals for the Ordered μ_j in Analysis With Four Sites (Left Interval), and Five Sites (Right Interval). These are 98% equal tail intervals with the medians, quartiles, and 5% and 95% points marked.

tion (1), across a range of y values, averaging with respect to the 40,000 posterior draws of all parameters. There are sharp differences between the case $s = 3$ and the others, but $s = 4$ and 5 produce rather similar density estimates. Increasing s beyond 5 makes little difference to the estimated density function. Corresponding informal comparisons of the cumulative distributions plotted against the empirical cdf of the signal data indicate close conformity of the data with cases $s = 4$ or 5, but apparent discrepancies with models based on fewer than four sites. Further graphical diagnostics based on inverting the predictive cdf and plotting against ordered data values provides a useful, if somewhat informal assessment of model adequacy. Such graphical diagnostics and comparisons agree with more formal analyses in the model allowing $q > 0$ with s specified as the maximum number of active sites, and using the $\text{Be}(q|s-1,1)$ prior for q, discussed earlier. (Similar results arise with a uniform prior.) Approximate posterior probabilities on the number h of active sites are as follows: at $s = 3$, $P(h = 3|\mathbf{y}, s = 3) \approx 1$; at $s = 4$, $P(h = 3|\mathbf{y}, s = 4) \approx .14$ and $P(h = 4|\mathbf{y}, s = 4) \approx .86$; at $s = 5$, $P(h = 4|\mathbf{y}, s = 5) \approx .49$ and $P(h = 5|\mathbf{y}, s = 5) \approx .51$; at $s = 6$, $P(h = 4|\mathbf{y}, s = 6) \approx .30$, $P(h = 5|\mathbf{y}, s = 6) \approx .43$, and $P(h = 6|\mathbf{y}, s = 6) \approx .27$. These summaries reflect the inherent uncertainties in estimating the number of active sites and the problem of overfitting as higher numbers are permitted. Though the prior for q marginally favors lower values of h, increasing s allows more sites and the model naturally tends to overfit, attempting to tailor itself to increasing minor features in the data, as is common in other varieties of mixture models. On the basis of parsimony, $h = 4$ is identified as a minimally acceptable number of active sites, and $h = 5$ seems relatively favored. Models with $s = h = 4$ and $s = h = 5$ are discussed further.

Figures 6 and 7 summarize the marginal posteriors for μ and π in each of the analyses with four and five active sites. Figure 6 displays approximate 98% equal-tailed posterior intervals for $\mu_1, \ldots, \mu_s$, for each of $h = s = 4$ and 5, with the levels ordered for identification; the vertical bars in each case run from the lower 1% to the upper 99% posterior quantile and have medians, quartiles, 5% and 95% points marked. At site index $j = 1$, for example, the two vertical

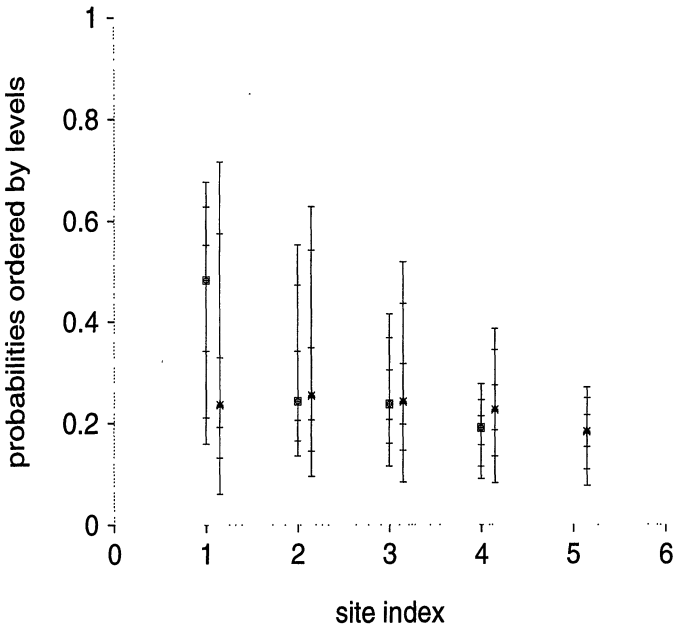

Figure 7. Summaries From Analyses of EPSP-A Data: At Each Site Index j, Posterior Intervals for the π_j, in Order of Increasing Values of the μ_j. As in Figure 6, the two analyses represented have four sites (left interval) and five sites (right interval).

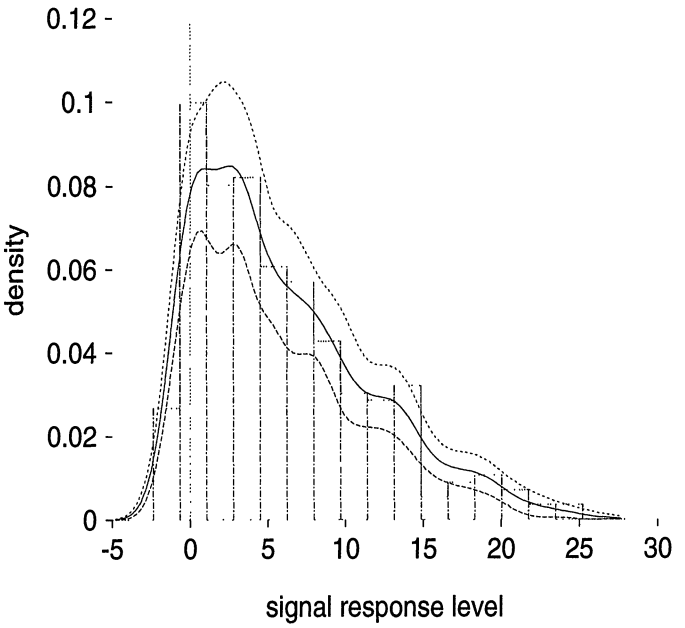

Figure 8. Summaries From Analysis of EPSP-A Data With s = h = 4 Active Sites: Data Histogram and the Predictive Density From the Analysis. Uncertainty about the predictive density is crudely indicated via pointwise bands representing ± 1.65 posterior standard deviations about the density estimate.

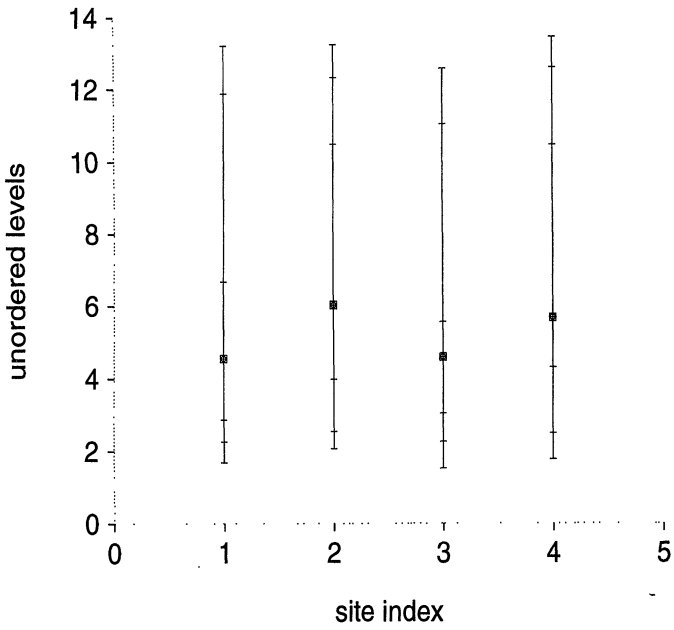

Figure 9. Summaries From Analysis of EPSP-A Data With s = h = 4 Active Sites: Posterior Intervals for the Unidentified Release Levels μ_j.

bars correspond to posterior intervals for μ_1, the smallest release level, from each of the two separate analyses. The vertical scale is meV. At $j = 5$, naturally there is only one interval, for μ_5 from the second analysis. There are apparent differences; analysis with just four sites naturally favors higher μ_j values at higher levels to accommodate the skewed upper tail of the signal data. In each analysis, the graphs indicate differences among the site release levels. This is supported by approximate posterior probabilities on equal/quantal values in the two analyses, which indicate strong support for at least three distinct values among the levels of active sites in each case. Write h^* for the number of distinct values. Then, in analysis with $s = h = 4$, the approximate posteriors are $P(h^* = 3|\mathbf{y}, h = 4) = .64$, and $P(h^* = 4|\mathbf{y}, h = 4) = .32$, whereas in analysis with $s = h = 5$ these are $P(h^* = 3|\mathbf{y}, h = 5) = .58$, $P(h^* = 4|\mathbf{y}, h = 5) = .28$, and $P(h^* = 5|\mathbf{y}, h = 5) = .03$. For release probabilities π_j in order of increasing μ_j, intervals appear in a similar display in Figure 7. Now the differences are less apparent, though still evident. Write k^* for the number of distinct values among the probabilities. Then, in analysis with $s = h = 4$, the approximate posteriors are $P(k^* = 2|\mathbf{y}, h = 4) = .53$ and $P(k^* = 3|\mathbf{y}, h = 4) = .37$, and when $s = 5$ these are $P(k^* = 2|\mathbf{y}, h = 5) = .45$, and $P(k^* = 3|\mathbf{y}, h = 5) = .30$, for example. This does suggest more disparity among the release levels than among the release chances, although there are differences and considerable uncertainty. What is clear is that a strict quantal structure with a single common release level and a single common release probability is quite untenable. This dataset is also typical of many in that there is the appearance of more disparity among the release levels than among release probabilities.

Further summaries of the analyses with fixed $s = h = 4$ appear in Figures 8–10. In Figure 8, the predictive density is superimposed on a data histogram, with some indications of uncertainty in terms of pointwise intervals representing ± 1.65 approximate posterior standard deviations about the density estimate. Figures 9 and 10 display, site by site, approximate 98% intervals, with quantiles marked as in earlier figures, but now for the *unidentified* μ_j and π_j. Recall that the μ_j share a common marginal posterior, as do the π_j. These figures indicate that the Monte Carlo output sum-

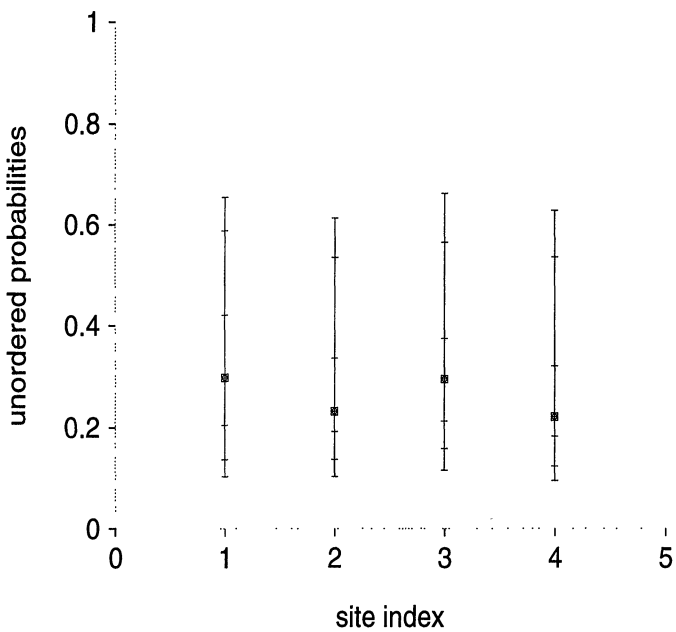

Figure 10. Summaries From Analysis of EPSP-A Data With s = h = 4 Active Sites: Posterior Intervals for the Unidentified Release Probabilities π_j.

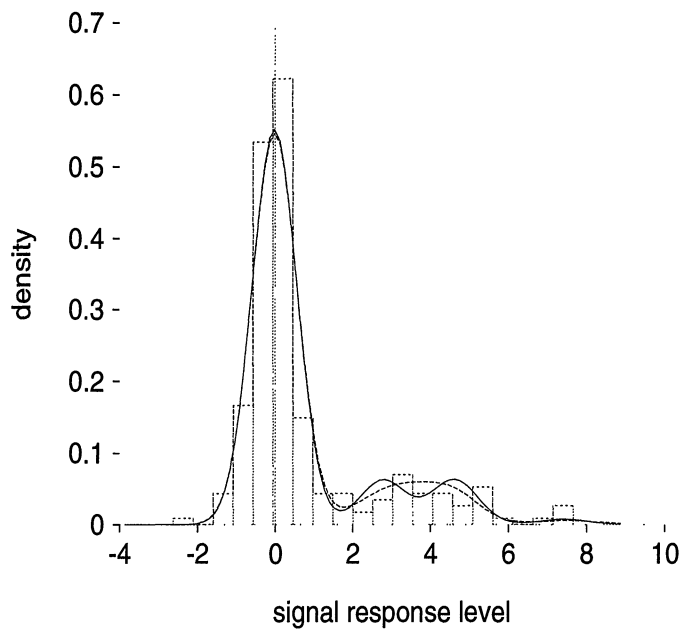

Figure 11. Summaries of Analyses of the EPSP-B1 Data: Posterior Predictive Densities From Two Separate Analysis With h = s = 2 and h = s = 3, Together With a Histogram of the Data. The density curves are from the model with h = 2 (full line) and h = 3 (dashed line).

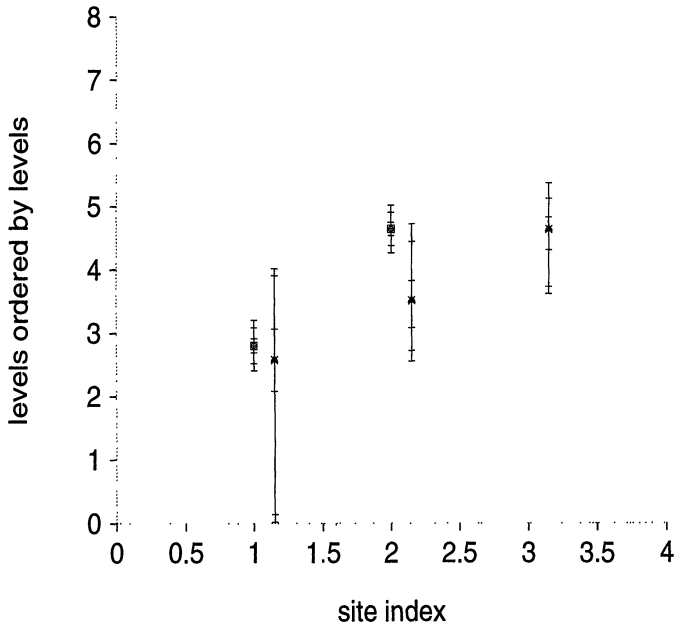

Figure 12. Summaries of Analyses of the EPSP-B1 Data: Posterior Intervals for the Ordered μ_j. At each j, the left interval corresponds to analysis with h = s = 2 sites; the right interval, to h = s = 3 sites.

marized is in close conformity to this theoretical exactness, acting as a confirmatory convergence diagnostic in this case.

Some comments about convergence of the Gibbs sampling algorithms from a practical perspective are in order. The predictive density function and posteriors for both ordered release levels and release probabilities displayed in the figures are stable over iterations, and in fact converge quite rapidly; the general features displayed appear after a few thousand, sometimes a few hundred iterations of the simulation algorithm. In this example, as in most real analyses, the true posterior margins are multimodal, reflecting the lack of identification of the unconstrained parameters; in the full parameter space there are multiple modes in the joint posterior. MCMC iterations tend to "stick" around a posterior mode, sometimes for a long while, and this is exacerbated when there are evidently distinct and separated release levels and/or probabilities, so that the modes are widely separated. Although the lack of convergence and lack of identification is "problematic," it works to an advantage here, as discussed earlier. Mapping to the ordered/identified parameters induces physical interpretation, and convergence is not such an issue, because the regions of unconstrained parameter space "explored" by the simulation are simply reflected under all parameter permutations. Also, the extent to which the margins in the unconstrained parameters differ is related to the extent of real differences in underlying site levels and/or chances; bigger differences induce more widely separated concentrations of posterior mass, and usually posterior modes, resulting in poor convergence for the unconstrained parameters.

5. ANALYSIS OF PAIRED-PULSE EPSP DATASETS

Figures 11–13 provide outputs from an analysis of the first of the paired-pulse datasets EPSP-B1 in Figure 3, in a format similar to that just described for the EPSP-A analyses. In analysis with $s = 4$ and the $\text{Be}(q|s-1,1)$ prior on q, approximate probabilities $P(h = 2|\mathbf{y}, s = 4) \approx .61$, $P(h = 3|\mathbf{y}, s = 4) \approx .24$, and $P(h = 4|\mathbf{y}, s = 4) \approx .16$ are found, suggesting just two or at most three active sites; the issues of overfitting and parsimony considerations discount the .16 probability on four active sites. Figures 12 and 13 display posterior intervals for the release levels and chances under models with fixed numbers of sites $h = s = 2$ and $h = s = 3$; recall that these intervals have medians, quartiles, and 1%, 5%, 95%, and 99% points marked.

Note the close similarity of predictive density functions in Figure 11, the main difference being that the model with two sites leads to more evident minor modes around 3 meV and 5 meV. From Figure 13, it appears that the two or three sites have rather low release probabilities around .07–.12. The data strongly support quantal structure in these probabilities, with $P(k^* = 1|\mathbf{y}, h = 2) = .94$ and $P(k^* = 1|\mathbf{y}, h = 3) = .84$. Release levels are likely distinct, with $P(h^* = 2|\mathbf{y}, h = 2) \approx 1$ and $P(h^* = 3|\mathbf{y}, h = 3) = .55$ and with estimated values near 2.5–3 meV and 4.5–5 meV for a two-site model and around 2–3 meV, 3–4 meV, and 4–5 meV for a three-site model. From Figure 12, note the significant uncertainties about release levels under the three-site model, partly due to the small sample size. This, coupled with small release probabilities, means that there are a very few, heavily influential observations really informing about site levels; in particular, the data around 7.5 meV in the histogram are influential in determining the sum of two or three of the μ_j. Figure 13 indicates relatively higher posterior precision in estimating release probabilities, with the exception of π_1 in the three-site model. In this latter analysis, although much posterior mass for π_1 is concentrated in

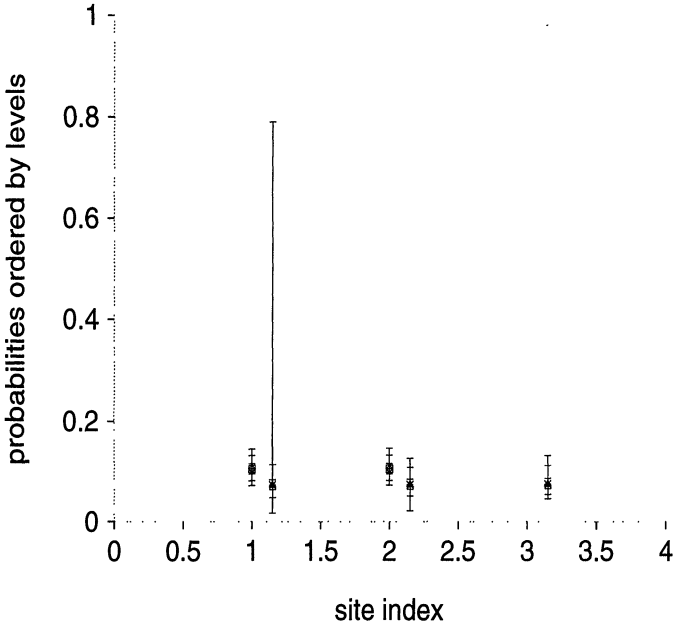

Figure 13. Summaries of Analyses of the EPSP-B1 Data: Posterior Intervals for the π_j in Order of Increasing μ_j. At each j, the left interval corresponds to analysis with h = s = 2 sites; the right interval, to h = s = 3 sites.

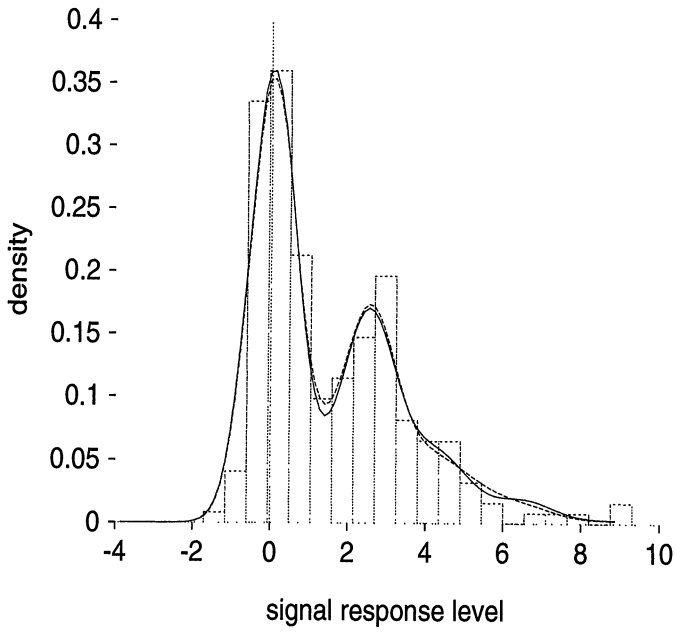

Figure 13. Summaries of Analyses of the EPSP-B2 Data: Posterior Predictive Densities From the Two Analyses Together With a Histogram of the Data. The density curves are from the model with h = 3 (full line) and h = 4 (dashed line).

the .07–.12 range, the marginal posterior is diffuse over a range of values consistent with the baseline uniform prior; this reflects the fact that the existence of a third site is dubious, related to the high degree of uncertainty about μ_1.

Figures 14–16 display similar graphical summaries of an analysis of the EPSP-B2 dataset, the second of the paired-pulse datasets. These are to be compared to the inferences for EPSP-B1 just described. It is apparent from the signal histogram that, working with the same synaptic tissue, there is more "action" in EPSP-B2 than in EPSP-B1 in the sense of higher EPSP levels. Neurologists are interested in whether this potentiation-induced activity is due to increased release probabilities or increased transmission levels, or both.

Analysis with $q > 0$ to assess the number of active sites indicates three or four sites. For a maximum $s = 4$, as in analysis of the first pulse, the approximate posteriors are $P(h = 3|\mathbf{y}, s = 4) = .32$ and $P(h = 4|\mathbf{y}, s = 4) = .68$. At a maximum of $s = 5$, these are $P(h = 3|\mathbf{y}, s = 5) = .13$, $P(h = 4|\mathbf{y}, s = 5) = .55$, and $P(h = 5|\mathbf{y}, s = 5) = .32$. This rather strongly indicates four active sites, particularly in view of the overfitting issue, which is an increased number relative to the first pulse, prepotentiation. The estimated predictive densities from three-, four-, and five-site models are almost indistinguishable; see Figure 14. Conditional on the likely four active sites, the marginal posterior on the number of distinct release levels have approximate masses of $P(h^* = 2, 3, 4|\mathbf{y}, h = s = 4) \approx .44$, .37, and .15; that for the number of distinct release probabilities is $P(k^* = 1, 2, 3|\mathbf{y}, h = s = 4) \approx .63$, .29, and .08. In summary, it appears reasonable to infer that the likely four active sites share a common release probability, but differ in release levels. This is again a consistent feature across a variety of experimental datasets. Release probabilities across sites are small; on a four-site model, the margins displayed in Figure 16 suggest that π_j lie in the region of .1–.2, with posterior medians around .14–.16.

Combining and comparing these inferences to the analysis of the EPSP-B1 data leads to the following conclusions about the nature of the synaptic junction being explored and the effects of the potentiation experiment. It seems plausible that there be four transmission sites. (Note that analysis based on $s = 5$ sites supports this conclusion.) Initially, sites are activated with a common but very low release probability, around .07–.1. In the first pulse only two or three sites are active, with differing transmission levels in the 3–5 meV range; if just two sites are active, then the levels are fairly precisely estimated at about 2.5–3 meV and 4.5 meV. Following potentiation, the second pulse, dataset is suggestive of the following effects. First, the site release probabilities are still constant across sites but have increased to around .14–.16 or higher, perhaps a doubling of the common release probability in the first pulse. This ties in with a leading hypothesis about the physiological effects of potentiation (see, e.g., Kullman and Nicoll 1992) that following potentiation, release sites are in a state of more immediate readiness to transmit and thus will be more likely to do so at the rapid repeat stimulation in a paired-pulse experiment. Indeed, the support for more active sites in the second pulse suggests the existence of four sites, that are simply not evident in the first pulse experiment; potentiation brought these sites "to life," and they are evident in the second pulse. The effects on transmission levels seem to work in the opposite direction. On a four-site model, the second pulse has release levels generally lower than those of the two or three sites evident in the first pulse. Although this is a tentative conclusion in view of the high levels of posterior uncertainty about the ordered release levels in the second pulse, it is also consistent with physiological theory suggesting that release sites, having transmitted on one stimulus occasion, require a waiting period before transmitting again to similar maximum levels; that is, a period of replenishment.

6. ADDITIONAL VARIANCE COMPONENTS: MODEL EXTENSION

Recent attention has focused on questions of additional variability in EPSP outcomes due to so-called *intrinsic variability* in release levels of individual synapses (Turner et al. 1995a,b): I describe this concept here and define an elaborated class of models incorporating it. This concept gives rise to model modifications in which components of the normal mixture representation have variances that increase with level, and this leads to considerable complications, both substantive and technical. The technical complications have to do with developing appropriate model extensions and associated MCMC techniques to analyze the resulting models; some brief development is mentioned here, with examples. The substantive complication is essentially that competing models with and without intrinsic variance components cannot be readily distinguished on the basis of the observed EPSP data alone; an observed data configuration

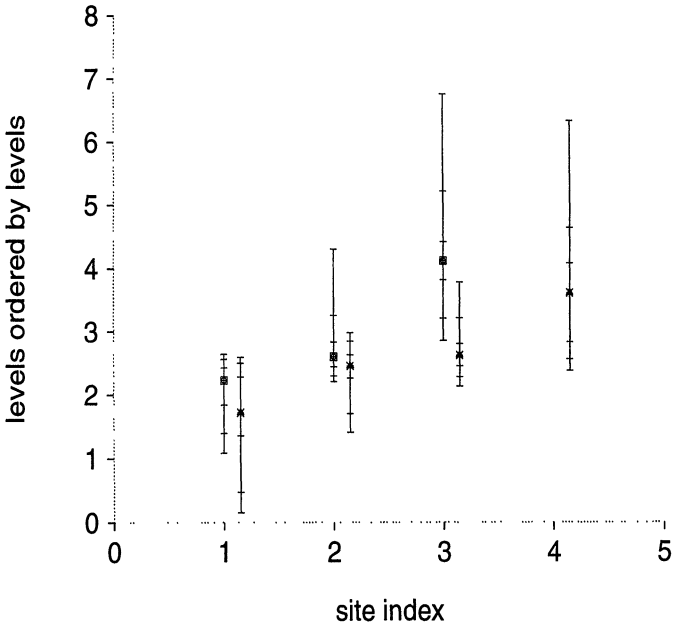

Figure 14. Summaries of Analyses of the EPSP-B2 Data: Posterior Intervals for the Ordered μ_j. At each site index, the left interval corresponds to analysis with h = s = 3 sites; the right interval, to h = s = 4 sites.

might arise from just one or two sites with significant intrinsic variance, or it might arise from a greater number of sites with low or zero intrinsic variance. In such cases, and especially when inferences about site characteristics are heavily dependent on the number of sites and levels of intrinsic variance, one is reliant on the opinions of expert neurophysiologists to judge between the models. Unfortunately, in its current state, the field is represented by widely varying expert opinions, from the one extreme of complete disregard of the notion of intrinsic variability to the other of belief in high levels of intrinsic variability as the norm. In some examples, including the EPSP-A dataset of this article, this issue is relatively benign, as inferences about release levels and probabilities are relatively insensitive to intrinsic variances. However, in other cases it is evidently highly relevant. Further collaborative research to refine knowledge of intrinsic variability effects is part of the current frontiers of the field. This section gives a short discussion of the notion and the current approach to modeling intrinsic variability.

Intrinsic variability refers to variation in levels of neurotransmitter release at specific sites. As developed previously, site j has a fixed release level μ_j, and although these levels may differ across sites, they are assumed to be fixed for the duration of the experiment under the controlled conditions. However, the mechanism of electrochemical transmission suggests that this may be an oversimplification. A site transmits by releasing a packet of (many) molecules of a chemical transmitter, and these molecules move across the synaptic "cleft" to the receiving cell. The induced potential response is proportional to the number of molecules received. So the assumption of fixed μ_j implies that (a) the number of molecules transmitted is constant across occasions, and (b) all ejected molecules are received by the postsynaptic cell. Each of these is questionable, and the issue has given rise to the notion of intrinsic variance; that is, variability in the site-specific level of release across occasions. Whatever the actual structure of variability, it is evidenced in EPSP datasets through mixture components with higher variance at higher release levels.

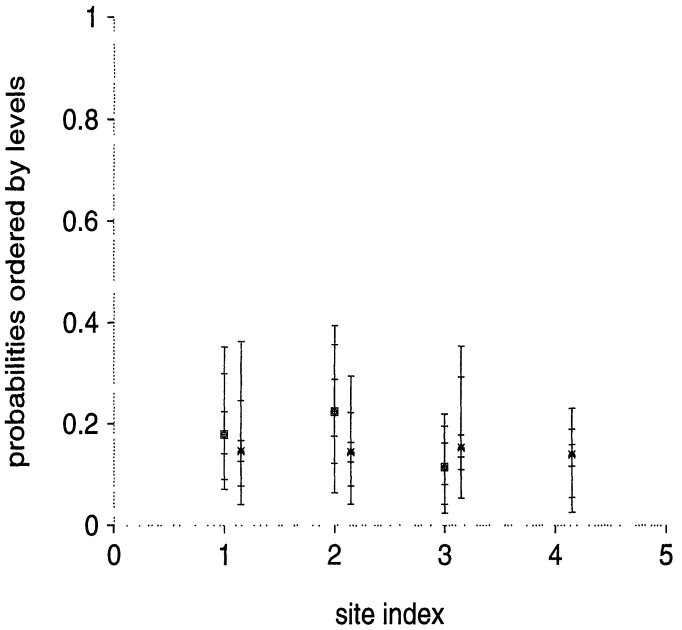

Figure 15. Summaries of Analyses of the EPSP-B2 Data: Posterior Intervals for the π_j in Order of Increasing μ_j. At each site index, the left interval corresponds to analysis with h = s = 3 sites; the right interval, to h = s = 4 sites.

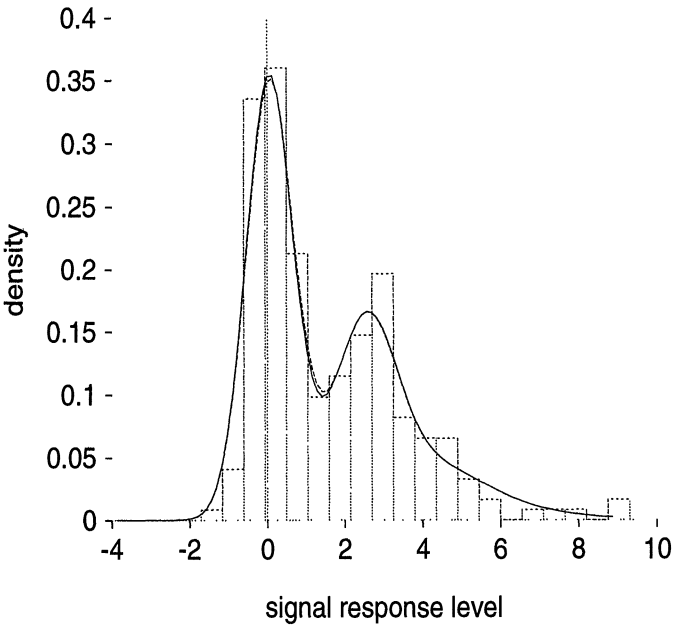

Figure 16. Summaries of Analyses of the EPSP-B2 Data, From Separate Analysis With h = s = 2 and h = s = 3, in the Model Extension to Incorporate Site-Specific Measures of Intrinsic Variability in Release Levels. The figure shows posterior predictive densities from the two analyses together with a histogram of the data. The density curves are from the model with h = 2 (full line) and h = 3 (dashed line).

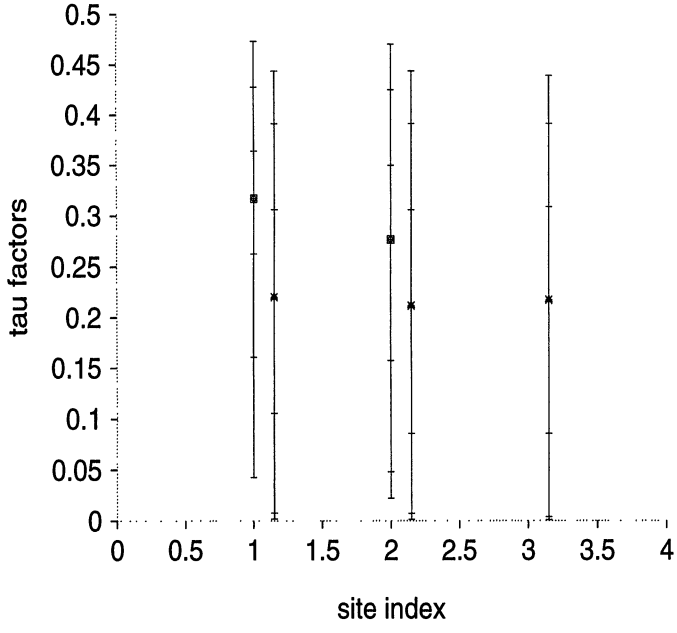

Figure 17. Summaries of Analyses of the EPSP-B2 Data, From Separate Analysis With h = s = 2 and h = s = 3, in the Model Extension to Incorporate Site-Specific Measures of Intrinsic Variability in Release Levels. Displayed are the usual posterior intervals for the τ_j factors with site indices representing increasing levels μ_j. At each site index, the left interval corresponds to analysis with h = s = 2 sites; the right interval, to h = s = 3 sites.

The basic extension of my models to admit the possibility of intrinsic variability involves remodeling the data y_i as coming from the conditional normal model

$$y_i \sim N(y_i|\theta_i, v) \quad \text{with} \quad \theta_i = m + \sum_{j=1}^{s} z_{ij}\gamma_{ij}, \gamma, \qquad (6)$$

where the γ_{ij} are new site- and case-specific release levels and v is the noise variance. Retaining independence across sites and trials, I assume that the γ_{ij} are distributed about an underlying expected release level μ_j—the same site-specific level as before, but now representing underlying average levels about which releases vary. Then the form of the distribution of the γ_{ij} about μ_j represents the intrinsic variability in induced responses due to variation in amounts of neurotransmitter released by site j and also due to variation in the success rate in moving the transmitter across the synaptic cleft. Various parametric forms might be considered; my preliminary work to date is based on exploration of models in which

$$p(\gamma_{ij}|\mu_j, \tau_j) \propto N(\gamma_{ij}|\mu_j, \tau_j^2\mu_j^2)I(0 < \gamma_{ij} < u) \qquad (7)$$

for all trials i and each site j; here u is the earlier specific upper bound on release levels. The new parameters $\tau_j > 0$ measure intrinsic variability of sites $j = 1, \ldots, s$; ignoring the truncation in (7), τ_j is an effective constant (though site-specific) coefficient of variation in release levels about the underlying μ_j. This model has been implemented, extending the prior structure detailed in Section 2 to incorporate the full set of site- and trial-specific release levels $\{\gamma_{ij}\}$ together with the new parameters $\tau_1, \ldots, \tau_s$. Note that as my model allows $\mu_j = 0$ for inactive sites, I use (7) only for $\mu_j > 0$; otherwise, $\mu_j = 0$ implies $\gamma_{ij} = 0$ for each $i = 1, \ldots, n$.

The effects of quantal variability are demonstrated by integrating the data density (6) with respect to (7). This is complicated due to the truncation to positive values; as an approximation, for the purposes of illustrating structure here, assume that this is not binding; that is, μ_j and τ_j are such that the mass of the basic normal distribution in (7) lies well withing the interval $(0, u)$. Then (6) and (7) combine and marginalize over γ_{ij} to give the approximation

$$y_i \sim N\left(y_i|m + \sum_{j=1}^{s} z_{ij}\mu_j, v + \sum_{j=1}^{s} z_{ij}\tau_j^2\mu_j^2\right). \qquad (8)$$

The mean here is as in the original formulation, Equation (1), with active sites contributing the expected release levels μ_j. But now the variance of y_i is not simply the noise variance v; it is inflated by adding in factors $\tau_j^2\mu_j^2$ for each of the active sites. Hence this is the feature relevant to modeling increased spread of EPSP data configurations at higher transmission levels.

It should be clear that this extended model can be managed computationally with direct extensions of the MCMC algorithms of Section 3 and Appendix B. I now turn to the full posterior $p(\boldsymbol{\mu}, \boldsymbol{\pi}, \mathbf{z}, \boldsymbol{\gamma}, \boldsymbol{\tau}, m, v, q, a, b|\mathbf{y})$, extending the original posterior in (5) to include the new quantities $\boldsymbol{\gamma} = \{\gamma_{ij}, i = 1, \ldots, n; j = 1, \ldots, s\}$ and $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_s\}$. There are difficulties in the posterior MCMC analysis due to the complicated form of (7) as a function of μ_j, and also due to the truncation of the basic normal model. These issues

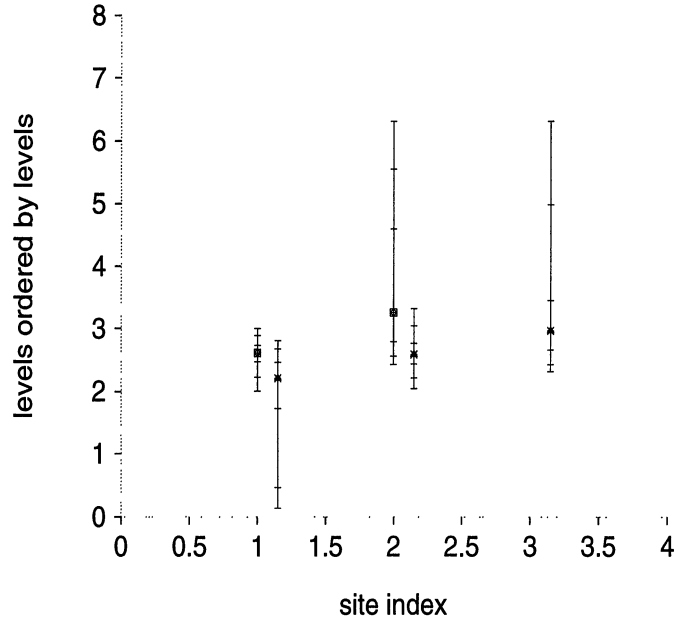

Figure 18. Summaries of Analyses of the EPSP-B2 Data, From Separate Analysis With h = s = 2 and h = s = 3, in the Model Extension to Incorporate Site-Specific Measures of Intrinsic Variability in Release Levels. Displayed are the usual posterior intervals for the ordered μ_j. The left interval corresponds to analysis with h = s = 2 sites; the right interval, to h = s = 3 sites.

destroy part of the nice conditionally conjugate sampling structure and are currently handled using some direct analytic approximations and Metropolis–Hasting accept/reject steps in the extended simulation analysis. (Details are very briefly reviewed in Appendix C.) It is beyond the scope of this article to fully develop the technical aspects of this here, but the discussion would be incomplete without raising the issue of intrinsic variability, currently in vogue in the field, and without providing some exploratory data analysis. Appendix C briefly details the major technical modifications to the analysis already spelled out in Section 3 and Appendix B. After further technical refinements and practical experience, I expect to report full modeling and technical details elsewhere. Here, for completeness, I briefly summarize results of analyses of some of dataset EPSP-B2 already discussed.

I take the τ_j to be a priori independent $\tau_j \sim U(\tau_j|0, .5)$ in the analysis, a maximum 50% level for the intrinsic coefficient of variation. Analysis of the EPSP-B2 data with a maximum of $s = 4$ active sites, and $q > 0$ to assess inactive sites, yields approximate probabilities $P(h = 2|\mathbf{y}, s = 4) = .19$, $P(h = 3|\mathbf{y}, s = 4) = .60$, and $P(h = 4|\mathbf{y}, s = 4) = .20$. Compare these to the earlier probabilities in the model with $\tau_j = 0$; that analysis led to a 68% probability of four active sites and a 32% probability of three active sites. So allowing intrinsic variability significantly reduces the most likely number of active sites. Figures 17–20 provide some summary graphics from two separate analysis with fixed numbers of sites $h = s = 2$ and $h = s = 3$, in a format similar to Figures 13–16. Note in Figure 17 the close similarity of predictive density functions from the two models.

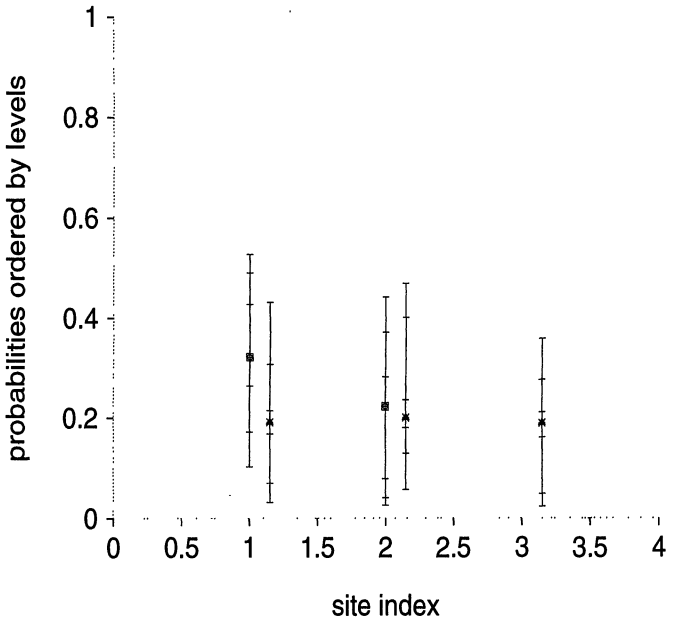

Figure 19. Summaries of Analyses of the EPSP-B2 Data, From Separate Analysis With h = s = 2 and h = s = 3, in the Model Extension to Incorporate Site-Specific Measures of Intrinsic Variability in Release Levels. Displayed are the usual posterior intervals for the π_j in order of increasing μ_j. The left interval corresponds to analysis with h = s = 2 sites; the right interval, to h = s = 3 sites.

Apart from impacting inferences on the number of active sites, the inferred average release levels μ_j are somewhat lower than in the earlier analysis with no intrinsic variance; (see Fig. 19). This is to be expected, as intrinsic variability now accounts for some of the higher observed levels of response. Figure 18 displays posterior intervals for the intrinsic variance parameters τ_j, in the same format as for release levels and chances in Figures 19 and 20. Note the inference of comparable levels across the likely two or three active sites, at appreciable levels of 10%–40% on the coefficient of variance scale though with high uncertainty.

Clearly, the two competing explanations—increased intrinsic variability at higher transmission levels versus more sites with fixed release levels—are in significant conflict in this example. The intrinsic variance model supports fewer sites, attributing the data at higher levels of transmission to inherently high levels of intrinsic variability. The earlier analysis, constrained to no intrinsic variance, needs more active sites to adequately cater for the observed data. Choosing between models involves committing to one of the extremes—yes or no in favor of intrinsic variability. Current physiological research, and a keener focus on the molecular levels of neurotransmission, may shed light on the issue. We note also there is no imperative to accept either model on the standard bases of simplicity and parsimony. Though an intrinsic variance model leads to fewer active sites, and hence is apparently a simpler description of the data-generating mechanism, it in no way represents a simpler model in terms of numbers of parameters and model complexity, as it requires many more latent γ_{ij} parameters to describe the data.

7. CONCLUDING COMMENTS

Besides the scientifically and statistically live issues arising in connection with investigations of intrinsic variability, several other areas are of current interest and are under initial investigation. As presented, our models make no explicit allowance for outliers in signal recordings, or for systematic features such as arise when synaptic tissue exhibits inhibitory as well as excitatory responses to stimuli. An example of the latter was discussed by West and Turner (1994), for example. Another related feature of some EPSP experiments is occasional (though rare) "blocking" of the responses of all or a subset of transmission sites, leading to more signal records than expected located around the zero level. There are obvious ways to extend the current models to allow for these kinds of events and considerations; one is to add mixture components explicitly representing such phenomena. Perhaps the most needed generalization is to relax the assumption of normally distributed experimental noise. This is currently in development, using noise models based on work in Bayesian density estimation following Escobar and West (1995) and West, Müller, and Escobar (1994). (See Cao and West 1996 for related developments.)

One beginning area of research involves issues of dependence in release characteristics over time and across sites. For the first issue, synaptic responses in general may be viewed as a stream where at each site the response may

depend on previous occurrences. Current analyses assume independence. The very nature of the paired-pulse potentiation experiments, however, explicitly recognizes dependence effects, in terms of both response/no response and response levels and the importance of time elapsed between successive stimuli generating the responses. Statistical developments of time series issues are of interest here to explore and, ultimately, model time variation in patterns of synaptic transmission. Interest in transmission dependencies also extends across sites (i.e., in possible "connections" between sites that may have physiological interpretation and importance), although serious inroads into study of this issue is some way in the future.

It is apparent that the technical aspects of simulation based analysis in the novel mixture models here bear further study, from a practical perspective. Investigating variations on the basic Gibbs sampling schemes that enhance and improve convergence characteristics is certainly desirable. It is possible that hybrid schemes based in part on the use of auxiliary variables (see, e.g., Besag and Green 1993) will prove useful here; the need for improved computational tools will be more strongly felt in more complex models incorporating nonnormal noise components. That said, the current approach described and illustrated here provides a useful basis for exploring and assessing EPSP data in attempts to more directly evaluate the structure and stochastic characteristics of synaptic transmission as it is currently viewed in the neurophysiological community. Current work on the applications side focuses on refining this basic framework and in exploring ranges of EPSP datasets to isolate and summarize the diversity of observed response mechanisms in mammalian central nervous systems.

APPENDIX A: COMPONENT PRIOR DISTRIBUTIONS

The Dirichlet structure of the component prior distribution $F(\cdot)$ in the prior modeling for nonzero site levels μ_j in Section 2.2.2 is detailed here. Write $f(\cdot)$ for the density of $F(\cdot)$. Recall that for some probability q, the μ_j are conditionally independent and drawn from

$$(\mu_j|F,q) \sim q\delta_0(\mu_j) + (1-q)f(\mu_j)$$

for $j = 1,\ldots,s$. Now $F(\cdot)$ is assigned a Dirichlet process prior with base measure given by $aU(\mu_j|0,u)$ for some precision $a > 0$ and the specified range $u > 0$. Marginalizing over F, the full joint prior for all s elements of $\boldsymbol{\mu}$ is now implicitly defined by

$$\begin{aligned} p(\boldsymbol{\mu}|q,a) &= \int p(\boldsymbol{\mu}|F,q)\,dP(F|a) \\ &= \int \left\{\prod_{j=1}^{s} p(\mu_j|F,q)\right\} dP(F|a) \\ &= \int \left\{\prod_{j=1}^{s} q\delta_0(\mu_j) + (1-q)f(\mu_j)\right\} dP(F|a). \end{aligned}$$

This is a rather complicated expression whose structure is most easily, and usefully, understood by exploring the corresponding conditional priors. For each $j = 1,\ldots,s$,

$$\begin{aligned} p(\mu_j|\boldsymbol{\mu}_{-j},q,a) &= \int p(\mu_j|F,\boldsymbol{\mu}_{-j},q)\,dP(F|a,\boldsymbol{\mu}_{-j}) \\ &= \int p(\mu_j|F,q)\,dP(F|a,\boldsymbol{\mu}_{-j}) \\ &= q\delta_0(\mu_j) + (1-q)E(dF(\mu_j)|a,\boldsymbol{\mu}_{-j}). \end{aligned}$$

Under the Dirichlet prior, the function $E(dF(\mu_j)|a,\boldsymbol{\mu}_{-j})$ is trivially deduced; write h_j for the number of nonzero elements of $\boldsymbol{\mu}_{-j}$ and N_j for the corresponding set of indices, $N_j = \{i|\mu_i > 0, i = 1,\ldots,s; i \neq j\}$. Then, using results of Ferguson (1973) or West et al. (1994) leads to

$$E(dF(\mu_j)|a,\boldsymbol{\mu}_{-j}) = r_j U(\mu_j|0,u) + (1-r_j)h_j^{-1}\sum_{i\in N_j}\delta_{\mu_i}(\mu_j),$$

where $r_j = a/(a+h_j)$. This yields the full joint prior $p(\boldsymbol{\mu}|q,a)$ defined via the set of conditionals displayed in Equation (3) of the text.

APPENDIX B: STRUCTURE OF CONDITIONAL POSTERIORS

More detail on the conditional posterior distributions underlying the MCMC sampling algorithm of Section 3 is provided here. Each of the seven conditional distributions is detailed, with comments about sampling at each step in the MCMC iterations.

B.1 Conditional Posteriors for Site Levels μ

For each j, there are the conditional prior (3) for μ_j and the associated conditional likelihood function

$$p(\mathbf{y}|\boldsymbol{\mu},\mathbf{z},m,v) = \prod_{i=1}^{n} N(y_i|m+\mathbf{z}_i'\boldsymbol{\mu},v).$$

Write $M_j = \sum_{i=1}^{n} z_{ij}$. Now $M_j = 0$ iff, $z_{ij} = 0$ for all i, so that the conditioning indicator vector $\mathbf{z}_i$ corresponds to no transmissions at all from the synaptic site labelled j. In this case $p(\mathbf{y}|\boldsymbol{\mu},\mathbf{z},m,v)$ does not depend on μ_j, and hence the conditional posterior for μ_j collapses to the prior (3). Otherwise, site j is identified as transmitting for at least some of the signal observations, and hence $M_j > 0$ and the likelihood reduces to a function of μ_j proportional to the normal density $N(\mu_j|m_j,v/M_j)$, where $m_j = M_j^{-1}\sum_{i=1}^{n} z_{ij}e_{ij}$ based on "residuals" $e_{ij} \stackrel{\text{def}}{=} y_i - m - \sum_{r=1,r\neq j}^{s} z_{ir}\mu_r$ for $i = 1,\ldots,n$.

Then, based on the conditional prior (3), it is deduced (for this case of $M_j > 0$) that

$$\begin{aligned} &p(\mu_j|\boldsymbol{\mu}_{-j},\mathbf{z},m,v,q,a,\mathbf{y}) \\ &\quad\propto q_{j0}\delta_0(\mu_j) + q_{jj}g_j(\mu_j) + \sum_{i\in N_j} q_{ji}\delta_{\mu_i}(\mu_j) \end{aligned}$$

with the following components:

- $q_{j0} = qN(0|m_j,v/M_j)$
- $q_{jj} = (1-q)r_jc_j/u$ with $c_j = \Phi((u-m_j)\sqrt{M_j}) - \Phi(-m_j\sqrt{M_j})$, and where $\Phi(\ )$ is the standard normal cdf
- $q_{ji} = (1-q)(1-r_j)h_j^{-1}N(\mu_i|m_j,v/M_j)$ for $i \in N_j$
- $g_j(\mu_j) = c_j^{-1}N(\mu_j|m_j,v/M_j)I(0 < \mu_j < u)$, the density of the normal distribution truncated to $0 < \mu_j < u$.

At each step of the MCMC iterations, sequence through $j = 1,\ldots,s$, at each step generating a new value of μ_j given the latest

sampled values of $\boldsymbol{\mu}_{-j}$ and all other conditioning quantities. For each j, this involves computing the $2 + h_j$ positive constants q_{ji}, $(i = 0, j; i \in N_j)$, then making a single multinomial draw with chances proportional to these q_{ji} to identify the selected mixture component. If $i = 0$ is selected, then $\mu_j = 0$; if $i > 0$ and $i \neq j$ is selected, then $\mu_j = \mu_i$; otherwise, $i = j$ and μ_j is drawn from $g_j(\cdot)$, simply the $N(\mu_j|m_j, v/M_j)$ truncated to $(0, u)$. Sampling efficiency is improved using variations on this basic routine involving configuration sampling, as introduced by MacEachern (1994) and developed by West et al. (1994).

B.2 Conditional Posteriors for Site Release Probabilities π

As a function of any single π_j, the posterior (5) trivially implies that

$$p(\pi_j|\boldsymbol{\pi}_{-j}, z, b) \propto \pi_j^{t_j}(1-\pi_j)^{n-t_j} p(\pi_j|\boldsymbol{\pi}_{-j}, b),$$

where $t_j = \sum_{i=1}^n z_{ij}$. Under the prior (4), this becomes the mixture

$$p(\pi_j|\boldsymbol{\pi}_{-j}, z, b) \propto s_{jj}\text{Be}(\pi_j|t_j+1, n-t_j+1) + \sum_{i=1, i\neq j}^{s} s_{ji}\delta_{\pi_i}(\pi_j),$$

where $\text{Be}(\cdot|\cdot,\cdot)$ represents the beta density and the s_{ji} are positive constants given by $s_{jj} = wt_j!(n-t_j)!/(n+1)!$ and $s_{ji} = (1-w)\pi_i^{t_j}(1-\pi_i)^{n-t_j}/(s-1)$ for each $i \neq j$.

At each step on the MCMC iterations, sequence through $j = 1, \ldots, s$, at each step generating a new value of π_j given the latest sampled values of $\boldsymbol{\pi}_{-j}$ and all other conditioning quantities. For each j, this involves computing the s positive constants s_{ji}, $(i = 1, \ldots, s)$, then making a single multinomial draw with chances proportional to these s_{ji} to identify the selected mixture component. If $i \neq j$ is selected, then $\pi_j = \pi_i$; otherwise, $i = j$ and π_j is drawn from the beta distribution $\text{Be}(\pi_j|t_j+1, n-t_j+1)$. Here again, configuration sampling improves efficiency.

B.3 Conditional Posteriors for Indicators $\mathbf{z}$

As a function of the 2^s vectors $\mathbf{z}_i$, (5) simplifies as $\prod_{i=1}^n p(\mathbf{z}_i|\boldsymbol{\mu}, \boldsymbol{\pi}, m, v, y_i)$, where

$$p(\mathbf{z}_i|\boldsymbol{\mu}, \boldsymbol{\pi}, m, v, y_i) \propto p(\mathbf{z}_i|\boldsymbol{\pi})p(y_i|\boldsymbol{\mu}, \mathbf{z}_i, m, v)$$

$$\propto \left\{\prod_{j=1}^{s} \pi_j^{z_{ij}}(1-\pi_j)^{1-z_{ij}}\right\} N(y_i|m + \mathbf{z}_i'\boldsymbol{\mu}, v).$$

These 2^s unnormalized probabilities are easily computed to provide a conditional multinomial posterior for the binary vector $\mathbf{z}_i$, and this is easily sampled. A complete draw of a new value for the full vector $\mathbf{z}$ simply involves running through independent simulations for each $i = 1, \ldots, n$.

B.4 Conditional Posterior for Systematic Bias m

The conditional likelihood function in m is a normal form, giving $p(m|\boldsymbol{\mu}, \mathbf{z}, v, \mathbf{y}) \propto p(m|v)\exp(-\sum_{i=1}^n (e_i - m)^2/(2v))$ based on residuals $e_i = y_i - \mathbf{z}_i'\boldsymbol{\mu}$; this reduces to a density proportional to $p(m|v)N(m|\bar{e}, v/n)$, which is normal as $p(m|v)$ is normal. This posterior is trivially sampled.

B.5 Conditional Posterior for Noise Variance v

The conditional likelihood function in v is a standard inverse gamma form, resulting in the posterior $p(v|\boldsymbol{\mu}, \mathbf{z}, m, \mathbf{y}) \propto p(v|m)v^{-n/2}\exp(-S/(2v))$, where $S = \sum_{i=1}^n (y_i - m - \mathbf{z}_i'\boldsymbol{\mu})^2$ for each i. Under the inverse gamma prior $p(v|m)$ from the noise data analysis, this conditional posterior is also inverse gamma, and so is easily sampled.

B.6 Conditional Posterior for Hyperparameter q

Knowledge of $\boldsymbol{\mu}$ implies knowledge of $h \leq s$, the number of nonzero elements in $\boldsymbol{\mu}$, and a binomial likelihood function for q. Under the specific beta prior $q \sim \text{Be}(q|s-1, 1)$, the resulting posterior is $(q|h) \sim \text{Be}(q|2s-h+1, h+1)$. Under an alternative uniform prior, this would change to $(q|h) \sim \text{Be}(q|s-h+1, h+1)$. Other beta priors yield appropriate beta posteriors.

B.7 Conditional Posterior for Hyperparameter a

The treatment here follows West (1992) and Escobar and West (1995). Assume a specified gamma prior $G(a|a_0, a_1)$ with density $p(a) \propto a^{a_0-1}e^{-a_1 a}$ for $a > 0$. Data augmentation ideas show that the relevant conditional posterior for a given $\boldsymbol{\mu}$ (and everything else) depends only on distinct, nonzero values among the elements of $\boldsymbol{\mu}$. So for any $\boldsymbol{\mu}$, write $h \leq s$ for the number of nonzero values and $h^* \leq h$ for the number of distinct nonzero values. Then the required conditional density $p(a|\boldsymbol{\mu})$ is the margin of a joint density $p(a, x|\boldsymbol{\mu})$ with the following conditionals:

- given x in (0, 1),

  $$(a|x, \boldsymbol{\mu}) \sim p_x G(a|a_0 + h^*, a_1 - \log(x)) + (1-p_x)G(a|a_0 + h^* - 1, a_1 - \log(x))$$

  where $p_x/(1-p_x) = (a_0 + h^* - 1)/\{h(a_1 - \log(x))\}$
- given $a > 0$, $(x|a, \boldsymbol{\mu}) \sim \text{Be}(x|a+1, h)$.

Although the corresponding density $p(a|\boldsymbol{\mu})$ is not directly manageable, the expression through conditionals is natural in the context of iterative posterior simulation analyses, as described in the aforementioned references. The parameter space is simply augmented with the latent variable x introduced to determine a joint posterior $p(a, x|\boldsymbol{\mu})$ with manageable conditionals. Write a_- for the value of a available from the latest iteration, so that the current values of h and h^* have been generated based on that value a_-. To sample a new value of a at the current iteration, generate a value of x from the beta distribution $p(x|a_-, \boldsymbol{\mu})$ and then sample from the gamma mixture $p(a|x, \boldsymbol{\mu})$.

B.8 Conditional Posterior for Hyperparameter b

This is treated essentially as a. Assume a specified gamma prior $b \sim G(b|b_0, b_1)$. Given a vector $\boldsymbol{\pi}$, write $s^* \leq s$ for the number of distinct values (all are nonzero under the assumed prior). Then values of b are iteratively sampled in a fashion similar to a, now based on linked conditional distributions as follows: For some w in (0, 1), the required conditional density $p(b|\boldsymbol{\pi})$ is the margin of a joint density $p(b, w|\boldsymbol{\pi})$ with the following conditionals:

- given w,

  $$(b|w, \boldsymbol{\pi}) \sim p_w G(b|b_0 + s^*, b_1 - \log(w)) + (1-p_w)G(b|b_0 + s^* - 1, b_1 - \log(w))$$

  where $p_w/(1-p_w) = (b_0 + s^* - 1)/\{s(b_1 - \log(w))\}$
- given $b > 0$, $(w|b, \boldsymbol{\pi}) \sim \text{Be}(w|b+1, s)$.

The sampling scheme generates a new value of b by augmenting the parameter space with w, in a way directly analogous to sampling a.

APPENDIX C: CONDITIONAL POSTERIORS WITH THE INTRINSIC VARIANCE MODIFICATION

This appendix briefly details the posterior structure and new components of the MCMC analysis required to incorporate the new sets of parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$ in the model extension for intrinsic

variability of Section 6. With reference to the original framework in Section 3, detailed in Appendix B, there are several changes to the simulation of conditional posteriors, and two additional steps. The set of components labelled 1–8 in Appendix B are modified and extended as follows.

C.1 Conditional Posteriors for Site Levels μ

The μ_j are sampled from conditional posteriors that are marginalized over γ. For each site index j in turn, there are the usual conditional prior 3 for μ_j and the associated conditional likelihood functions from 8; the latter are now more complicated as functions of μ_j and result in posteriors that are not directly simulated. Instead, analytic normal approximations are used as Metropolis–Hasting (M–H) proposal distributions and an M–H independence chain step is embedded within the overall MCMC loop. Write m_j for the "current" value of μ_j in the iterations, and fix the variances of 8 by plugging in m_j for μ_j, for each j. This leads to approximate likelihood components that are normal, as in the original scheme of Appendix B, though now with differing but known variances. Step 1 of Appendix B can now be completed as written to obtain a candidate draw for μ_j, and the associated M–H accept/reject ratio is easily computed and used to accept or reject the candidate value. Sequencing through $j = 1, \ldots, k$ using this strategy produces a new vector $\boldsymbol{\mu}$.

C.2 Conditional Posteriors for γ and τ

The new MCMC steps draw release levels γ_{ij} and then intrinsic coefficients of variance τ_j from appropriate conditional posteriors. First, we sample the elements of γ. For fixed i and j, note that, conditional on all other parameters and the observed data, the conditional posterior for the individual release level γ_{ij} is proportional to the product of Steps 6 and 7. If $z_{ij} = 0$, then this reduces to the prior 7. Otherwise, 6 and 7 trivially combine to give a normal posterior multiplied by the truncation indicator in 7. Either way, γ_{ij} is trivially sampled from the resulting truncated normal distribution.

Second, we sample the elements of $\boldsymbol{\tau}$. Suppose that the prior has the τ_j independent with common density $g(\tau_j)$ (a uniform prior in the current implementation). Then, conditional on γ and $\boldsymbol{\mu}$, the prior combines with the conditional likelihood components arising from 7 to yield independent conditional posteriors for the τ_j. In case $\mu_j = 0$, 7 is vacuous and the posterior for τ_j is simply the prior, so easily sampled. Otherwise,

$$p(\tau_j|\gamma, \boldsymbol{\mu}) \propto g(\tau_j)c(\tau_j)^{-n} \exp(-G_j/2\tau_j^2)/\tau_j^n,$$

where $G_j = \sum_{i=1}^{n}(\gamma_{ij}/\mu_j - 1)^2$ and $c(\tau_j)$ is the probability on $(0, u)$ under the $N(\cdot|\mu_j, \tau_j^2\mu_j^2)$ distribution. This has the form of an inverse gamma density multiplied by $g(\tau_j)c(\tau_j)^{-n}$, suggesting a M–H/independence chain step for sampling. Draw from the inverse gamma so determined as an M–H proposal density, and then accept/reject according to the resulting, and easily computed, M–H ratio.

C.3 Conditional Posteriors for Site Release Probabilities π

This follows the original model formulation in Appendix B. Steps 3–5 inclusive in Appendix B remain essentially the same, but with levels γ_{ij} replacing the μ_j in the means of the conditional normal distributions for the y_i throughout. The remaining steps 6–8 are unchanged.

[Received September 1994. Revised August 1996.]

REFERENCES

Besag, J., and Green, P. J. (1993), "Spatial Statistics and Bayesian Computation," *Journal of the Royal Statistical Society*, Ser. B, 55, 25–37.

Clements, J. D. (1991), "Quantal Synaptic Transmission?," *Nature*, 353, 396.

Cao, G., and West, M. (1996), "Bayesian Analysis of Mixtures of Mixtures," *Biometrics*, 52, 221–227.

Escobar, D. M., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Kullman, D. M. (1989), "Application of the Expectation-Maximization Algorithm to Quantal Analysis of Postsynaptic Potentials," *Journal of Neuroscience Methods*, 30, 231–245.

——— (1992), "Quantal Analysis Using Maximum Entropy Noise Deconvolution," *Journal of Neuroscience Methods*, 44, 47–57.

Kullman, D. M., and Nicoll, R. A. (1992), "Long-Term Potentiation is Associated With Increases in Quantal Content and Quantal Amplitude," *Nature*, 357, 240–244.

Ling, L., and Tolhurst, D. J. (1983), "Recovering the Parameters of Finite Mixtures of Normal Distributions From a Noisy Record: An Empirical Comparison of Differing Estimating Procedures," *Journal of Neuroscience Methods*, 8, 309–333.

MacEachern, S. N. (1994), "Estimating Normal Means With a Conjugate-Style Dirichlet Process Prior," *Communications in Statistics, Part B—Simulation and Computation*, 23, 727–741.

Martin, A. R. (1966), "Quantal Nature of Synaptic Transmission," *Physiology Review*, 46, 51–66.

Redman, S. (1990), "Quantal Analysis of Synaptic Potentials in Neurons of the Central Nervous System," *Physiology Review*, 70, 165–198.

Smith, A. F. M., and Roberts, G. O. (1993), in "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Ser. B, 55, 3–23.

Tierney, L. J. (1994), "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22, 1701–1728.

Titterington, D. M., Smith, A. F. M., and Makov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*, Chichester, U.K.: Wiley.

Turner, D. A. (1987), "Identification of Components of Evoked Composite EPSP's in CA1 Hippocampal Pyramidal Neurons," *Neuroscience Abstracts*, 13, 155.

Turner, D. A., Chen, Y., Isaac, J., West, M., and Wheal, H. V. (1995a), "Heterogeneity Between Non-NMDA Synaptic Sites in Paired-Pulse Plasticity of CA1 Pyramidal Neurons in the Hippocampus," *Journal of Physiology*, in press.

Turner, D. A., Isaac, J., Chen, Y., Stockley, E., and Wheal, H. V. (1995b), "Analysis of Dendritic Synaptic Sites in Hippocampal CA1 Pyramidal Cells—Assessment of Variability and Non-Uniformity," in *Excitatory Amino Acids and Synaptic Transmission*, ed. H. V. Wheal and A. M. Thomson, London: Academic Press.

Turner, D. A., and Schlieckert, M. (1990), "Data Acquisition and Analysis System for Intracellular Neuronal Signals," *Journal of Neuroscience Methods*, 35, 241–251.

Turner, D. A., and West, M. (1993), "Statistical Analysis of Mixtures Applied to Postsynaptic Potential Fluctuations," *Journal of Neuroscience Methods*, 47, 1–23.

Turner, D. A., and Wheal, H. V. (1991), "Excitatory Synaptic Potentials in Kainic Acid Denervated Rat CA1 Pyramidal Neurons," *Journal of Neuroscience*, 11, 2786–2794.

Walmsley, B., Edwards, F. R., and Tracey, D. J. (1987), "The Probabilistic Nature of Synaptic Transmission at a Mammalian Excitatory Central Synapse," *Journal of Neuroscience*, 7, 1037–1046.

——— (1988), "Nonuniform Release Probabilities Underlie Quantal Synaptic Transmission at a Mammalian Excitatory Central Synapse," *Journal of Neuroscience*, 60, 889–908.

West, M. (1992), "Hyperparameter Estimation in Dirichlet Process Mixture Models," ISDS Discussion Paper 92-A03, Duke University.

West, M., and Cao, G. (1993), "Assessing Mechanisms of Neural Synaptic

Activity," in *Bayesian Statistics in Science and Technology: Case Studies*, eds. C. Gatsonis, J. Hodges, R. Kass, and N. Singpurwalla, New York: Springer-Verlag.

West, M., Müller, P., and Escobar, M. D. (1994), "Hierarchical Priors and Mixture Models, With Applications in Regression and Density Estimation," in *Aspects of Uncertainty: A Tribute to D. V. Lindley*, eds. A. F. M. Smith and P. Freeman, Oxford, U.K.: Oxford University Press.

West, M., and Turner, D. A. (1994), "Deconvolution of Mixtures in Analysis of Neural Synaptic Transmission," *The Statistician*, 43, 31–43.

Wong, K., and Redman, S. J. (1980), "The Recovery of a Random Variable From a Noisy Record With Application to the Study of Fluctuations in Synaptic Potentials," *Journal of Neuroscience Methods*, 2, 389–409.