Testing the Fit of a Quantal Model of Neurotransmission

Anders C. Greenwood,* Elliot M. Landaw,# and Thomas H. Brown*§

*Department of Cellular and Molecular Physiology, Yale University, New Haven, Connecticut 06510; #Department of Biomathematics, University of California, Los Angeles, California 90095; and §Department of Psychology, Yale University, New Haven, Connecticut 06510, USA

ABSTRACT Many studies of synaptic transmission have assumed a parametric model to estimate the mean quantal content and size or the effect upon them of manipulations such as the induction of long-term potentiation. Classical tests of fit usually assume that model parameters have been selected independently of the data. Therefore, their use is problematic after parameters have been estimated. We hypothesized that Monte Carlo (MC) simulations of a quantal model could provide a table of parameter-independent critical values with which to test the fit after parameter estimation, emulating Lilliefors's tests. However, when we tested this hypothesis within a conventional quantal model, the empirical distributions of two conventional goodness-of-fit statistics were affected by the values of the quantal parameters, falsifying the hypothesis. Notably, the tests' critical values increased when the combined variances of the noise and quantal-size distributions were reduced, increasing the distinctness of quantal peaks. Our results support two conclusions. First, tests that use a predetermined critical value to assess the fit of a quantal model after parameter estimation may operate at a differing unknown level of significance for each experiment. Second, a MC test enables a valid assessment of the fit of a quantal model after parameter estimation.

INTRODUCTION

Much recent interest in quantal analysis was aroused by its use in studies of hippocampal long-term potentiation (reviewed in Stevens, 1993; Bekkers, 1994; Jack et al., 1994) and other forms of synaptic plasticity. Recent applications of quantal analysis to synaptic plasticity have extended from frequency facilitation in lobster neuromuscular junction (Worden et al., 1997) to long-term depression in rat neocortex (Torii et al., 1997). Many of these studies have relied on a parametric model, a practice that raises two questions. First, the question of "model discrimination": does a given model fit the data significantly better than simpler alternatives, and not significantly worse than more complex alternatives? Second, the question of "goodness of fit": does a given model fit the data as closely as expected, given that the data deviate from model predictions only because of sampling error? A model that has been selected by careful discrimination among alternatives need not fit the data closely enough to avoid rejection by a test of fit, as recently noted in the quantal-analysis literature (Greenwood, 1995; Stricker et al., 1996). The theory behind 1) model discrimination and 2) testing goodness of fit is commonly treated in textbooks (Stuart and Ord, 1991). Much recent computational work in quantal analysis falls in the first category of model discrimination (Smith, 1993; Stricker et al., 1994, 1996; Greenwood, 1995) and has antecedents in other fields (Wilks, 1938: Akaike, 1974; Horn, 1987; McLachlan, 1987). The present study addresses the second issue of testing goodness of fit.

Classically, a test of fit tests the hypothesis that the model is "true" (correct in form and parameter values). If the model has been fitted to the data and parameters have been estimated, the hypothesis is only that the model is correct in form. If information from the data has been included in the model, the model is "data dependent," and testing goodness of fit is a complex problem. When parameters are estimated, this data-dependency problem is called the "prior estimation problem." The problem of evaluating the fit of data-dependent models has not been adequately addressed in the quantal analysis literature. For example, the χ^2 test after correction for prior estimation of parameters is highly dependent the exact way in which the data are arranged in a histogram ("binning"), unless advanced techniques are used (Nikullin and Greenwood, 1996).

Therefore, we used the conventional Kolmogorov and Cramer–von Mises test statistics, which are calculated without binning the data. To assess goodness of fit after parameter estimation, these statistics must be compared to empirical critical values, as the classical values apply only to data-independent models. Examining these statistics in the context of the original quantal model of transmission (Del Castillo and Katz, 1954), Monte Carlo (MC) simulations, and maximum likelihood estimation, we considered two methods of obtaining empirical critical values. First, we considered using MC-generated critical-value tables in emulation of Lilliefors's and Srinivasan's tests for the normal, uniform, and exponential distributions (Lilliefors, 1967, 1969; Srinivasan, 1970; Mason and Bell, 1986). Second, we considered using MC simulations to obtain a unique critical value for each data set (Press et al., 1992; Raastad and Lipowski, 1996). This second procedure, which we call the "MC test of fit," involves more computation than the first

Received for publication 10 October 1997 and in final form 5 January 1999.

Address reprint requests to Dr. Anders C. Greenwood, Department of Neurosciences, University of New Mexico, Albuquerque, NM 87131. Tel.: 505-272-0620; Fax: 505-272-8082; E-mail: agreenwo@unm.edu.

0006-3495/99/04/1847/09 $2.00

approach, which would involve a one-time batch of simulations to generate critical-value tables. Therefore, we intended to compute tables of critical values for testing the fit of quantal models, after verifying that consistent critical values were obtained for a variety of simulated quantal parameters. Instead, we found that these critical values actually depended on the parameters in a quantal model. This result undermines the idea of a critical-value table, although we do report some rules of thumb. Moreover, our results suggest that how closely a correctly formulated quantal model fits a specific data set after estimation depends on how distinctly quantal the data are. This hitherto unreported effect may even confound the χ^2 test but causes no problem in the MC test of fit, as its critical values are specific to each data set.

MATERIALS AND METHODS

The Poisson model of quantal transmission

This conventional model assumes that packets of transmitter are released with an equal and small probability from a large number of sites by a stable ("stationary") process that is independent of the electrical response amplitude of each quantum ("quantal size"). The mean number of quanta in an event is termed the quantal content (m). As the response to a packet of transmitter may show some variability (Bekkers et al., 1990; Clements et al., 1992; Faber et al., 1992; Kruk et al., 1997, but see Edwards et al., 1990; Liao et al., 1992; Kullmann, 1993), the quantal size was described in terms of its mean (q) and coefficient of variation (CV_q). Other conventional assumptions included linear summation of quanta, independent additive noise with zero mean, and mutually independent responses (reviewed in Stevens, 1993). For simplicity, we assumed that the distributions of noise and quantal size were Gaussian. We defined the signal-to-noise ratio (Q/N) as q over the noise standard deviation. In simulations of the Poisson release model, the number of quanta (k) was limited to k_{max}, defined such that the probability $P(k > k_{max}) < 0.001$.

Monte Carlo simulations

Batch simulations were controlled by a custom-written program (in Precision-Visuals Wave language) and C library functions. Many independent identically distributed sets of 200 simulated response amplitudes were generated. In the simulation studies we explored a range of values of m, CV_q, and Q/N, while restricting the sample size to an experimentally motivated worst-case figure of 200 for two reasons. First, parameter values are likely to change over long experiments or with short inter-stimulus intervals. Second, some plasticity factors might diffuse out of the cell into the pipette after a period of whole-cell recording. For example, before the tetanus in careful experiments on mossy-fiber long-term potentiation (LTP), 200–400 responses were typically collected at 0.2–0.25 Hz, and as few as 200 responses in a stable epoch were selected for analysis (Xiang et al., 1994).

Maximum likelihood estimation

Given sufficient data, ML estimation is usually an optimal method in the sense that parameter estimates are unbiased and approach optimal precision (Rao, 1970; Stuart and Ord, 1991). What constitutes sufficient data is a matter of noise and sample size, and it depends on the nature and complexity of the data-generating mechanism. For a particular choice of parameters, the probability density is calculated for each response amplitude. These densities are multiplied to obtain the likelihood of the data. The ML estimate is the set of parameter values that maximize the likelihood. To analyze data that may adhere to classical quantal models, we wrote a program that differs from others that use the expectation maximization algorithm (Dempster et al., 1977; Kullmann, 1989; Stricker et al., 1994), in that it maximizes the explicitly calculated likelihood (see also Smith, 1993). This program was written in FORTRAN77 and uses IMSL library functions.

Testing goodness of fit

Problems with the χ^2 test in quantal analysis

The basic χ^2 test assumes that the model and the bin boundaries are data-independent. When model parameters have been estimated, the degrees of freedom of the test are commonly corrected by subtracting the number of estimated parameters. However, this correction is only valid if parameters have been estimated from bin occupancies, an inefficient method of estimation (Stuart and Ord, 1991; Nikullin and Greenwood, 1996). Furthermore, binning of data sets that are limited in sample size by the need for stationarity (Brown et al., 1976; Malinow, 1991) may lead to low expected bin occupancy and test results that depend on binning choices. Thus, use of the χ^2 test is typically problematic in quantal analysis. We developed a MC test of fit, using Kolmogorov–Smirnov (KS) and Cramer–von Mises (CvM) statistics to avoid such problems.

Kolmogorov–Smirnov and Cramer–von Mises tests

The KS and CvM statistics quantify the deviation between model and data in usefully different ways, as explained below (Stephens, 1974). These statistics compare the theoretical cumulative distribution function (cdf) for the model to the empirical or "observed" cdf of the data. The KS statistic is approximately the square root of n multiplied by the maximum deviation between the observed and theoretical cdf (Birnbaum, 1952; Press et al., 1992). In contrast, the CvM statistic measures the average squared deviation between these cdfs (von Mises, 1964). The CvM and KS statistics complement each other in that the CvM statistic is more sensitive to the tails of the cdfs. Also, our experience has been that the KS test's sensitivity to one large deviation makes it sensitive to quantal discrepancies from a unimodal model after ML estimation on quantal data (data not shown), despite its reported insensitivity to such discrepancies in the absence of prior estimation (Stratford et al., 1997). We used MC methods to approximate unknown post-estimation critical values for these statistics.

MC test of fit

In general, the MC test of fit tests the ability of the model formulation (without predetermined parameters) to account for the data. More specifically, this MC procedure tests whether the difference between the data and the fitted ("data-derived") model is consistent with the difference remaining after the same model formulation has been fitted to simulated data from the data-derived model (see Fig. 6). To apply the MC test of fit to data from the hippocampal mossy-fiber synapse in rat brain slices (Greenwood, 1995), we typically used an empirical reference distribution from 400 or 500 simulations of the estimated ML model, with each simulation consisting of 200 points. Fit statistics were obtained for each simulated data set and its ML model. These fit statistics were placed in ascending order to get an empirical cdf. As an estimate of the $\alpha_{0.05}$ critical value, the 95th percentile sample in this empirical cdf was compared to the experimental test statistic. For a batch of 500 simulation-derived fit statistics, the 25th largest statistic was an estimate of the $\alpha_{0.05}$ critical value. As underestimating this critical value would have artificially increased the test's power, we sometimes constructed a 95% confidence interval for the critical value (Stuart and Ord, 1991) and used its upper bound in place of the estimated critical value. Thus, the chance of underestimation was ~0.025.

RESULTS

Classical tests of fit and prior estimation

An exposition of the prior estimation effect (Figs. 1 and 2) will introduce the new results (shown in Figs. 3–5). Together with the new results, this exposition may motivate adoption of the MC test of fit (illustrated in Fig. 6). Fig. 1 shows histograms of 200 samples that were simulated using a model with Poisson release and Gaussian quantal size ($m = 1$, $q = 1$, $CV_q = 0.3$, noise $\sigma = 0.2$). In Fig. 1 *A*, this model's density function and its quantal components are shown along with a blindly chosen histogram. The quantal peaks become less distinct with increasing multiples of q (*dotted lines*), as the quantal variance is compounded and added to the noise variance. We quantified the goodness of the true model's fit with the Kolmogorov–Smirnov (KS) and Cramer–von Mises (CvM) statistics, obtaining p-values of 0.89 and 0.68 from the classical tables. In Fig. 1 *B*, the true density is shown with a hand-picked deviant histogram that has unusually few simulated failures of transmission. Both tests rejected the true model ($p < 0.01$), illustrating that the power to reject incorrect models comes with a probability of incorrectly rejecting the true model (significance level $\alpha = 0.01$).

Fig. 1 *C* illustrates a typical situation in quantal analysis, in which parameters are estimated in a model that is then justified by statistical argument. In Fig. 1 *C*, the histogram from Fig. 1 *B* is shown with the density of the standard model after maximum likelihood (ML) estimation. We estimated $m = 1.15$, $q = 1.05$, and $CV_q = 0.253$ from the data and assumed the simulated noise variance. The agreement between the estimated density and the first histogram peak in Fig. 1 *C* shows that the estimation of $m = 1.15$ accounted well for the simulated transmission failures. The classical KS and CvM tests failed to reject this ML model ($p_{CvM} = 0.65$, $p_{KS} = 0.8$), illustrating the fact that prior estimation decreases the chance that a test will reject a correctly formulated model, increasing the significance of any rejection. Thus, prior estimation requires some correction analogous to decreasing the degrees of freedom in a χ^2 test. For example, without such a correction an estimated model

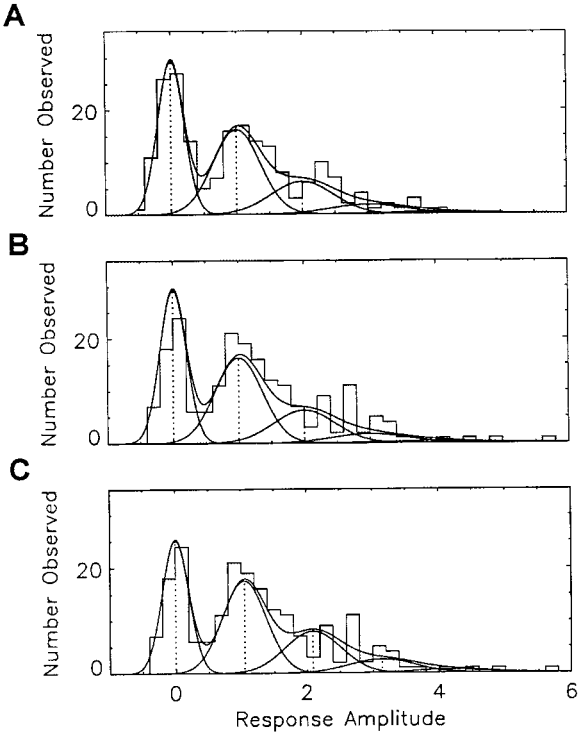

FIGURE 1 Prior estimation and significance level in a test of fit. *A* and *B* illustrate the concept of significance level in a classical test of fit. A model synapse with Poisson release and Gaussian quantal size was simulated to produce the histograms of 200 response amplitudes that are shown in *A* and *B*. The parameters were $m = 1$, $q = 1$, $CV_q = 0.3$, and noise $\sigma = 0.2$. The histogram in *A* was blindly selected, and the histogram in *B* was selected because it deviated from the simulated model. The smooth density function of the simulated model is shown with both histograms. The quantal components are indicated by Gaussian curves and vertical dotted lines. The closeness of fit of the simulated model was quantified with Kolmogorov–Smirnov (KS) and Cramer–von Mises (CvM) statistics. The fit to the blindly selected data set of *A* was not rejected ($p_{KS} = 0.89$ and $p_{CvM} = 0.68$). Using classical critical values for $\alpha = 0.01$, both tests rejected the fit of the true model to the unusual data shown in *B*. *B* and *C* show that estimating parameters before the application of a classical test decreases the chance of incorrectly rejecting a correctly formulated model, which generally implies a decrease in the power of the test to reject incorrectly formulated models. The same histogram of simulated data is shown in *B* and *C*, but in *C* the smooth density function of the best-fitting model is shown instead of the simulated model's density function. The parameters of the best-fitting model were $m = 1.15$, $q = 1.05$, and $CV_q = 0.253$. The correct noise value was assumed ($\sigma = 0.2$). As a consequence of the prior estimation of parameters from the same data, the fit of this model could not be rejected by the classical KS and CvM tests ($p_{KS} = 0.80$ and $p_{CvM} = 0.65$).

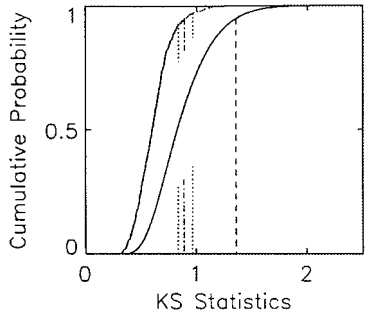

FIGURE 2 Prior estimation shifts the cumulative distribution of KS statistics to the left of the classical distribution. The model that was introduced in Fig. 1 ($m = 1$, $q = 1$, $CV_q = 0.3$, and noise $\sigma = 0.2$) was simulated to produce 400 sets of 200 response amplitudes. From each data set we obtained the KS statistic for the ML model after estimation of m, q, and CV_q. The KS statistics were placed in ascending order, and their ranks in the sequence were divided by 400 for plotting in an empirical cumulative distribution (*left curve*). The classical distribution of KS statistics for the fit of a correct independent model is indicated by the solid right curve. Dashed lines (interrupted for clarity) mark the $\alpha_{0.05}$ critical values in the theoretical and the empirical distributions. The dashed-dotted lines that bracket the empirical critical value (0.89) represent a 95% confidence interval for the critical value of the true KS distribution for this model after parameter estimation. The prior estimation of three parameters in this model is thus shown to shift the critical value from the classical value (~1.36) to a much smaller value.

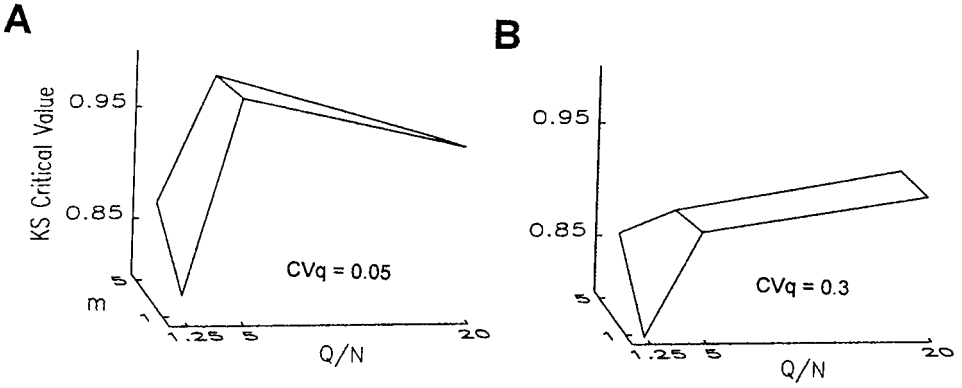

FIGURE 3 Critical values for the MC test of fit (at $\alpha = 0.05$) were found to have a complex dependence on the parameters of a model synapse with Poisson release and Gaussian quantal size. The magnitude of the $\alpha_{0.05}$ critical value reflects the amount of residual discrepancy between the data and the model that will be exceeded by 5% of correctly formulated models after estimation. Empirical KS critical values that were obtained for models after estimation were smaller than the classical value for the fit of data-independent models (~1.36). MC KS critical values were obtained and plotted for CV_q = 0.05 (*A*) and $CV_q = 0.3$ (*B*) over the domain of the other parameters (*m*: 1, 5; *Q*/*N*: 1.25, 5, 20). For each synapse, 400 sets of 200 response amplitudes were simulated. MC critical values were obtained from the KS fit statistics for the 400 best-fitting models. The average half-width of the 95% confidence intervals for these critical values was 0.05. The critical values increased when *Q*/*N* was increased from 1.25 to 5 in *A* ($CV_q = 0.05$) and from 5 to 20 in *B* ($CV_q = 0.3$). Thus, lower levels of combined variance from the noise and quantal-size distributions appeared to correlate with larger critical values after parameter estimation (see also Fig. 4).

including simple binomial release may appear to fit data that were simulated with nonstationary and/or nonuniform binomial parameters (Brown et al., 1976). This example concerns the "power" of a test of fit to reject a false model, which depends on the details of the true and false models and on the test's significance level. The present work focuses on significance to obtain results with general implications regarding power.

Quantifying the effect of prior estimation

In the example in Fig. 1, *B* and *C*, prior estimation improved the goodness of fit from $p < 0.01$ to $p \approx 0.7$, as measured by the classical KS and CvM tests. Expanding on the same example ($m = 1$, $q = 1$, $CV_q = 0.3$, noise $\sigma = 0.2$), Fig. 2 shows the effect of prior estimation on a test-statistic distribution and its $\alpha_{0.05}$ critical value. We performed ML estimation on each of 400 data sets and calculated KS and CvM statistics for each data set's ML model. The curve on the left in Fig. 2 shows the discrete cumulative distribution of the KS statistics (*dots*, unresolvable at high density). Plotted as a solid curve is an approximate classical distribution of the KS statistic (Press et al., 1992), from which the post-estimation distribution has been shifted to the left. The vertical dotted-dashed line (interrupted for clarity) indicates the empirical 95th percentile value (0.89), an estimate of the post-estimation $\alpha_{0.05}$ critical value. The bracketing dotted lines (interrupted for clarity) indicate this value's 95% confidence interval, which does not include the classical $\alpha_{0.05}$ critical value (~1.36, *dashed line*). In fact, none of the post-estimation KS statistics exceed this value. Thus none of the fits would be rejected by the classical test. Presumably, the classical test would also lack power to reject incorrectly formulated models. Largely similar results were obtained for the CvM statistics (data not shown).

Prior estimation calls for a model-specific test of fit

If the parameters of the simulated model in Figs. 1 and 2 had been estimated from experimental data, the empirical KS critical value in Fig. 2 (*left dotted-dashed line*) could be used to test the fit of this model. Not anticipating that the effects of prior estimation would depend on model parameters, we hypothesized that an empirical critical value such as this one would be appropriate to test the fit of the standard model after *m*, *q*, and CV_q had been estimated from 200 data points simulated with any set of parameters in the standard model. To test this hypothesis, we conducted a preliminary MC study. As the inherent variability of the quantum is a controversial issue in the experimental realm (Bekkers et al., 1990; Edwards et al., 1990; Liao et al., 1992) and in the realm of biophysical modeling (Clements et al., 1992; Faber et al., 1992; Kullmann, 1993; Kruk et al., 1997), CV_q was simulated to be 0.05 or 0.3, spanning much of the range of CV_q that is contested in this literature. Under each of these conditions, *m* was simulated to be 1 or 5, and the quantal signal-to-noise ratio was $Q/N = 1.25$, 5, or 20. For each combination of parameters, we obtained empirical $\alpha_{0.05}$ critical values from 400 KS and CvM statistics.

The KS critical values are shown over the domain of the simulation parameters in Fig. 3. Fig. 3, *A* and *B*, shows results for $CV_q = 0.05$ and 0.3, respectively. The classical $\alpha_{0.05}$ critical value for the fit of data-independent models is ~1.36, well above the surface. For clarity, the MC uncertainty in the critical-value estimates is not shown in Fig. 3.

However, for the case in which $CV_q = 0.3$, $m = 1$, $Q/N = 5$, this uncertainty is shown in Fig. 2 by the 95% confidence interval (*interrupted dotted lines*). The average half-width of these 95% confidence intervals for the critical values in Fig. 3 is 0.05. Thus, the range of critical values in Fig. 3 undermines the hypothesis that these values would all be MC estimates of one critical value. Instead, as Q/N increases from 1.25, the critical values increase to a higher level at $Q/N = 5$ in Fig. 3 *A* ($CV_q = 0.05$) and at $Q/N = 20$ in Fig. 3 *B* ($CV_q = 0.3$). Similar results were obtained with the CvM statistic (data not shown). This dependence on Q/N and CV_q suggests the new hypothesis that the variances of the noise and quantal-size distributions have a cumulative effect on the critical values after parameter estimation. We next subjected this hypothesis to closer examination.

Closeness of fit depends on Q/N and CV_q

Fig. 4 *A* shows KS critical values (for $\alpha = 0.05$) that were obtained from 11,200 simulations of the standard model synapse at each of seven different levels of simulated recording noise ($Q/N = 1.25, 1.98, 3.15, 5, 7.94, 12.6, 20$). The other parameters were $m = 1$ and $CV_q = 0.3$. Although m also contributes to the variance that reduces the distinctness of quantal peaks, it was not varied because its special role in our standard model could complicate the results (as in Fig. 5). The error bars represent approximate 95% confidence intervals for the expected reproducibility of each critical-value estimate. Fig. 4 *A* shows a sigmoidal dependence of the KS critical value on Q/N. The critical value is shown to be significantly larger for values of $Q/N \geq 3.15$ than for values of $Q/N \leq 1.98$. As a check of the practical significance of these differences, we determined the effect of using the critical value for $Q/N = 1.25$ to test the fit of estimated ML models to the data that had been simulated with $Q/N = 20$. Test statistics from 13% (instead of 5%) of these models exceeded this inappropriate critical value, leading to 160% too many rejections. Largely similar results were obtained for the CvM statistic (data not shown).

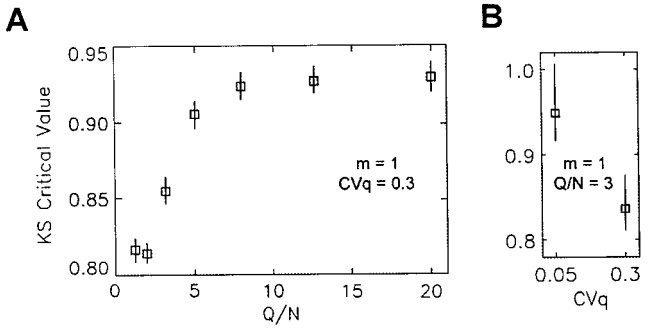

FIGURE 4 The KS critical value for the MC test of fit (at $\alpha = 0.05$) was found to have a sigmoidal dependence on Q/N (*A*) and decreasing CV_q increased the critical value, much as if Q/N had increased (*B*). These results mean that a greater discrepancy between the data and the model tended to remain when the model was fit to data with more clearly resolved quantal structure. (*A*) A model synapse with Poisson release and Gaussian quantal size was simulated to produce 11,200 sets of 200 response amplitudes at each of seven levels of simulated recording noise ($Q/N = 1.25, 1.98, 3.15, 5, 7.94, 12.6, 20$). The other parameters were $m = 1$ and $CV_q = 0.3$. For each data set a KS statistic was obtained for the fit of the best-fitting model. The $\alpha_{0.05}$ critical value from each distribution of KS statistics was plotted against Q/N. The error bars represent approximate 95% confidence intervals. The end points of this curve are (Q/N, KS critical value) = (1.25, 0.82) and (20, 0.93), whereas the KS critical value for a data-independent model is ~1.36. (*B*) Each critical value was obtained from 800 sets of 200 simulated response amplitudes. Critical values (*squares*) and error bars that represent approximate 95% confidence intervals are shown for $CV_q = 0.05$ and 0.3. For parameters ($Q/N = 3$; $m = 1$) that correspond to a point near the bottom of the region of maximum slope in *A*, a decrease in CV_q from 0.3 to 0.05 (moving from right to left in *B*) had the expected effect of increasing the $\alpha_{0.05}$ critical value. Thus, lower levels of combined variance from the noise and quantal-size distributions correlated with larger critical values after parameter estimation.

The region of maximum slope in Fig. 4 *A* is near the Q/N value of 5, for which the critical value increases as one looks from Fig. 3 *A* ($CV_q = 0.3$) to Fig. 3 *B* ($CV_q = 0.05$). To show a near-maximum effect of reducing CV_q from 0.3 to 0.05 in Fig. 4 *B*, we obtained MC critical values for the standard model with $m = 1$ and $Q/N = 3$ (near the bottom of the maximum-slope region in Fig. 4 *A*). The KS critical value was larger for $CV_q = 0.05$ than for $CV_q = 0.3$ (Fig. 4 *B*; error bars reflect ~95% confidence intervals). This result is consistent with the effect of increasing Q/N from a value of 3 in Fig. 4 *A*. As one would expect from the critical-value plateau at high values of Q/N in Fig. 4 *A*, reducing CV_q from 0.3 to 0.05 did not have a significant effect on the critical values for $m = 1$ and $Q/N = 20$ (data not shown). Thus, lower cumulative levels of variance in the noise and quantal-size distributions correlated with larger critical values after parameter estimation. As the quantal variance is compounded in multiquantal responses, we do not expect this cumulative effect to be strictly additive, and the details will depend on the distribution of quantal content. The general point is that more distinct quantal peaks correlated with larger critical values in this study. Similar results were observed for the CvM statistic (data not shown).

An implication of these results emerges from comparison of the range of KS critical values (0.82–0.93) in Fig. 4 *A* to the Lilliefors $\alpha_{0.05}$ KS critical values for a normal model with an estimated variance of 1.333 ($n > 100$; Mason and Bell, 1986) and an estimated mean and variance of 0.886 ($n > 30$; Lilliefors, 1967). For small Q/N in Fig. 4 *A*, the prior estimation of three parameters (m, q, and CV_q) reduced the KS critical value from its classical value of ~1.36 to values lower than the Lilliefors critical value for two estimated parameters, as seems appropriate. When Q/N was increased in the simulations, the post-estimation KS critical value increased in excess of this Lilliefors critical value, as if fewer than two parameters had been estimated. Although the different models involved preclude strict comparison, these observations suggest that distinct quantal structure may limit the loss of effective degrees of freedom that is caused by parameter estimation.

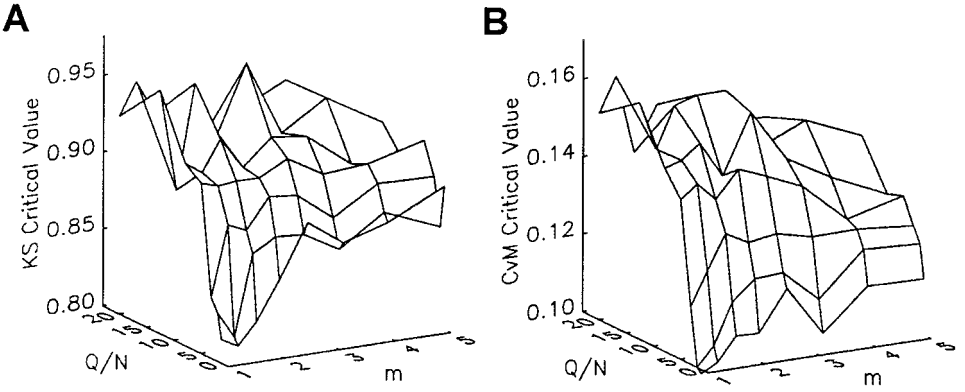

FIGURE 5 The effects of m and Q/N on critical values for the MC test of fit. Critical values for the MC test of fit increased with increasing Q/N, over a domain of m values that extended from 1 to 5. For low values of Q/N, the critical values also increased with increasing m. KS (A) and CvM (B) critical values are plotted over the simulated domain of Q/N and m. Each critical value was derived from 800 simulations of a set of 200 response amplitudes. The parameters were m = 1, 1.3, 1.63, 2.04, 2.55, 3.19, 3.99, 5; Q/N = 1.25, 1.98, 3.15, 5, 7.94, 12.6, 20; CV_q = 0.3. The average half-widths of the 95% confidence intervals for the KS and CvM critical values were 0.03 and 0.01, respectively, but the proximity of neighboring points in the surfaces suggests that the effect of sampling error is often much smaller. Note that the CvM critical values have proportionately less uncertainty than the KS critical values.

Complex dependence of critical values on *m* and *Q/N*

To examine how the critical values depended on m and Q/N, we used CV_q = 0.3 in simulations of a standard model synapse in which m was varied among eight values (1, 1.3, 1.63, 2.04, 2.55, 3.19, 3.99, 5) and Q/N was varied among seven values (1.25, 1.98, 3.15, 5, 7.94, 12.6, 20). For each parameter combination, we obtained KS and CvM critical values from 800 ML models fitted to simulated data sets (200 points each). These critical values are plotted in Fig. 5, A and B, respectively, showing that the critical values increased with increasing Q/N for all simulated values of m. Under noisy conditions and especially for $Q/N \leq 1.98$, increasing m also increased the critical values. Leaving aside possibly model-specific details, Fig. 5 shows that the dependence of the critical values on quantal parameters can be unpredictable and complex. This result argues for the use of a MC test instead of a critical-value table, when the fit of a quantal model is tested after estimation. However, our results do suggest some conservative, rule-of-thumb critical values for rejection or acceptance of a fit, in that no estimate of the $\alpha_{0.05}$ KS critical value fell above 0.95 or below 0.8 after estimation of m, q, and CV_q in our standard quantal model (see Figs. 3–5). The corresponding values for the CvM statistic were 0.16 and 0.1.

DISCUSSION

In quantal analysis, data from single experiments are often used both to estimate a model's parameters and to test its fit. However, classical tests of fit (including the χ^2, KS, and CvM tests) assume a data-independent model. In the χ^2 test, corrections for prior estimation are typically somewhat arbitrary or require that the parameters be estimated from binned data (Stuart and Ord, 1991; Nikullin and Greenwood, 1996). Therefore, we examined the prior-estimation problem in the context of KS and CvM statistics and a simple quantal model (Del Castillo and Katz, 1954). In this context, we found that the problem cannot have a solution that employs a single MC-generated critical value after the estimation of a specific group of parameters from a variety of data—essentially a quantal version of Lilliefors's and Srinivasan's tests for the uniform, normal, and exponential distributions (Lilliefors, 1967 and 1969; Srinivasan, 1970; Mason and Bell, 1986). Instead, we found that the appropriate critical values depended on model parameters, such as the variance in quantal size and the noise variance, which affect how distinctly the data are quantized. Although we report rules of thumb for interpreting test statistics in specific circumstances, the MC test of fit is a general solution.

This report is the first published study of the application of the MC test of fit to a quantal model. A similar MC approach using the ML value as a goodness-of-fit statistic was briefly described in a recent experimental study involving a nonquantal model (Raastad and Lipowski, 1996). In contrast, the present computational study used KS and CvM statistics because their distributions are known for data-independent models, permitting our examination of the effects of prior estimation. The ML value lacks this advantage and cannot easily be interpreted to test goodness of fit with tables or rules of thumb, because it depends very strongly on the data.

Advantages of the MC test of fit

The MC test of fit is an approach to the prior-estimation problem that can be easily used with any estimation method and any fit statistic that is calculated without binning the data. In a MC test of a model's fit to experimental data after estimation, critical values are not obtained from a standard distribution of reference statistics. Instead, as shown in Fig. 6, they are taken from a new distribution of statistics that

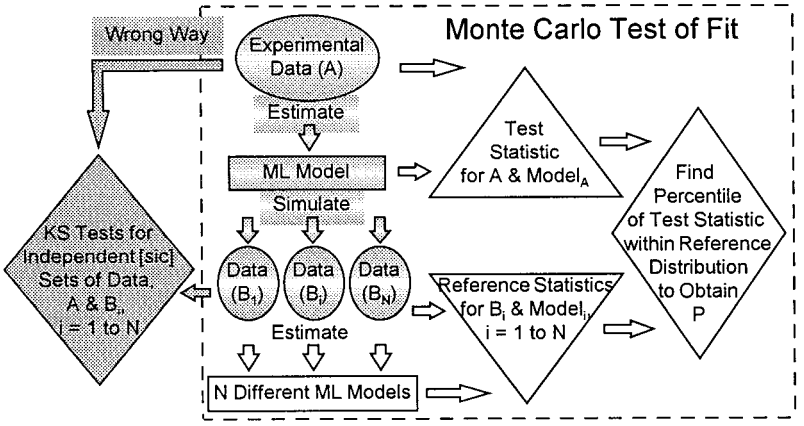

FIGURE 6 Schematic outline of the Monte Carlo test of fit (within *dashed rectangle*) and a simpler flawed alternative (*fully shaded elements on the left*). Elements unique to the MC test of fit are not shaded. Shared elements of the two tests are partly shaded. Arrows indicate information flow. Through estimation, information flows down from the experimental data to the ML model. Through simulation, information flows down from the ML model to a series of data sets that are independent of each other but have a special relation to the experimental data because they were generated from its ML model. This special relation leads to a flaw in a published alternative to the MC test of fit (Atwood and Tse, 1988). In this other test, the N simulated data sets are compared with the experimental data, using the KS test for two independent [sic] data sets, where "[sic]" marks the flaw. Returning to the flow of the MC test of fit, estimation on the N simulated data sets generates N different ML models. The first set of rightward arrows indicates the pairwise comparison of data sets with their ML models to generate the experimental test statistic (*top triangle*) and the reference statistic distribution (*bottom triangle*), which are brought together to evaluate the fit of the model to the experimental data (*final arrows leading to diamond*).

quantify the fit of models that have been fitted to sets of simulated data. These data are generated by simulating the model with parameters estimated from the experimental data. When models have been fitted to the simulated data sets, each "simulation-derived" parameter in these fitted models is distributed about the corresponding parameter that was estimated from the experimental data. These simulation-derived parameter distributions reflect the uncertainty of the data-derived parameters and complement other estimates of parameter uncertainty that involve more theory (McLachlan, 1978; Smith et al., 1991) or resampling (Stricker et al., 1994). Such estimates of parameter uncertainty will be more meaningful when a model has passed the MC test of fit.

Concerns regarding the MC test of fit

Concerns may remain about possible pitfalls in the use of the MC test of fit with quantal models. For example, the model may be correctly formulated, but the parameter estimates may be inaccurate, resulting in misleading simulations. In this case, the MC critical values may be different from those that would be obtained if the true parameters could be simulated. Two steps can be taken to address this concern. First, one can examine the scatter in the simulation-derived parameter estimates, as an indication of the likely accuracy of the estimates from the experimental data under the null hypothesis that the model is correctly formulated. Second, one can compare the magnitude of this scatter to the corresponding range of MC critical values that arises from the parameter dependence of the critical values. In a related study of our standard model, we found that 95% of parameter estimates typically fell within 10% of the true values, close enough that the parameter dependence of the MC critical values would be unlikely to cause errors (Greenwood, 1995). In this study's worst case, when CV_q was much smaller than Q/N ($CV_q = 0.05$, $Q/N = 1.25$, $m =$ 1 or 5), CV_q was typically overestimated by a factor of 2–4. However, comparison of the KS critical values for $CV_q =$ 0.05 and 0.3 in Fig. 3 suggests that even this overestimation would not effect the critical values when $Q/N = 1.25$. Nonetheless, parameter confidence limits will vary for different models (Stricker et al., 1994), as the parameter dependence of MC critical values probably will. Therefore, when the MC test of fit is applied to models that differ radically from our standard model, parameter uncertainty and critical-value sensitivity to parameters should be compared for a reasonable range of parameter values. Concerns regarding the random origin of the MC test's critical values were addressed in Materials and Methods (MC Test of Fit).

A simpler MC test of fit is flawed

We also considered a previously reported MC test of fit, in which data sets were simulated from an estimated model and compared to the experimental data with the KS test for two data sets (Atwood and Tse, 1988). The KS test for two data sets assumes, however, that the two sets are independently sampled from the same model. The dependence of the simulated model on the experimental data violated this assumption, as shown in Fig. 6. A MC study confirmed that the distribution of KS statistics that quantified the disparity between the experimental and simulated data sets was shifted to the left of the classical distribution (data not shown), increasing the test's actual significance level and

reducing its power to reject a fit at the nominal significance level.

Why tests of fit after estimation depend on quantal parameters

In fitting quantal data, a closer fit may be obtained by adjusting parameters such as σ and CV_q that describe the variance of continuously distributed variables. Distinct quantal peaks in the distribution constrain such parameters to small values, however, effectively reducing the number of estimated parameters. In more general terms, the parameter dependence of fit-testing critical values after estimation may arise in part from the extent to which variability from the continuous random processes in the model can account for variations that actually arose from the discrete random component of the synaptic mechanism. Our results are consistent with this view.

Other practical implications

Addressing an obvious point, our results confirmed that the KS and CvM fit statistics were smaller after ML estimation than would be expected for a data-independent model. It is also obvious that the number of estimated parameters affects how data-dependent a model is. However, it is less obvious that a model's data dependence varies with the ML-versus-entropy weighting factor in the maximum entropy noise deconvolution approach (Kullmann, 1992; Kullmann and Nicoll, 1992). In this context, it is worth noting that a constant critical-value criterion does not amount to a constant criterion for goodness of fit when the data dependence of a model varies. To apply a constant criterion for goodness of fit, the tolerance for discrepancy between data and model must be reduced as the model's data dependence is increased.

Inspired by the tests of Lilliefors and Srinivasan, we wanted to generate MC critical-value tables for a test of fit involving less computation than the MC test of fit. However, the parameter dependence of the critical values after estimation argues against such a test; its actual significance would vary with the quantal parameters of differing synapses. Nonetheless, our results suggested a rule of thumb for testing the fit of ML-estimated models that resemble our standard model. A CvM statistic that exceeded 0.16 would warrent rejection of the fit in all cases that we examined, whereas in none of these cases would a CvM statistic below 0.1 warrant rejection. The corresponding rule-of-thumb critical values for the KS statistic were 0.95 and 0.8. Such rules of thumb notwithstanding, the MC test of fit is a valuable step to include in model-based quantal analysis, providing support for results obtained in the areas of model discrimination, parameter estimation, and hypothesis testing.

This work was supported by National Cancer Institute grant CA16042, Office of Naval Research grant N00014-90-J-4136, National Institutes of Health grants MH50213 and MH56190, and a National Science Foundation graduate fellowship.

REFERENCES

Akaike, H. 1974. A new look at statistical model identification. *IEEE Trans. Automatic Control.* 19:716–723.

Atwood, H. L., and F. W. Tse. 1988. Changes in binomial parameters of quantal release at crustacean motor axon terminals during presynaptic inhibition. *J. Physiol. (Lond.).* 402:177–193.

Bekkers, J. M. 1994. Quantal analysis of synaptic transmission in the central nervous system. *Curr. Opin. Neurobiol.* 4:360–365.

Bekkers, J. M., G. B. Richerson, and C. F. Stevens. 1990. Origin of variability in quantal size in cultured hippocampal neurons and hippocampal slices. *Proc. Natl. Acad. Sci. USA.* 87:5359–5362.

Birnbaum, Z. W. 1952. Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. *J. Am. Statist. Assoc.* 47: 425–441.

Brown, T. H., D. H. Perkel, and M. W. Feldman. 1976. Evoked neurotransmitter release: statistical effects of nonuniformity and nonstationarity. *Proc. Natl. Acad. Sci. USA.* 73:2913–2917.

Clements, J. D., R. A. J. Lester, G. Tong, C. E. Jahr, and G. L. Westbrook. 1992. The time course of glutamate in the synaptic cleft. *Science.* 258:1498–1501.

Del Castillo, J., and B. Katz. 1954. Quantal components of the end-plate potential. *J. Physiol. (Lond.).* 124:560–573.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm (with Discussion). *Proc. R. Statist. Soc. Ser. B.* 39:1–38.

Edwards, F. A., A. Konnerth, and B. Sakmann. Quantal analysis of inhibitory synaptic transmission in the dentate gyrus of rat hippocampal slices: a patch clamp study. 1990. *J. Physiol. (Lond.).* 430:213–249.

Faber, D. S., W. S. Young, P. Legendre, and H. Korn. 1992. Intrinsic quantal variability due to stochastic properties of receptor-transmitter interactions. *Science.* 258:1494–1498.

Greenwood, A. C. 1995. Quantal analysis: parameter estimation and hypothesis testing. Ph.D. dissertation, Yale University.

Horn, R. 1987. Statistical methods for model discrimination: applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. J.* 51:255–263.

Jack, J. J., A. U. Larkman, G. Major, and K. J. Stratford. 1994. Quantal analysis of the synaptic excitation of CA1 hippocampal pyramidal cells. *Adv. Second Messenger Phosphoprotein Res.* 29:275–299.

Kruk, P. J., H. Korn, and D. S. Faber. 1997. The effects of geometrical parameters on synaptic transmission: a Monte Carlo simulation study. *Biophys. J.* 73:2874–2890.

Kullmann, D. M. 1989. Application of the expectation-maximization algorithm to quantal analysis of postsynaptic potentials. *J. Neurosci. Methods* 30:231–245.

Kullmann, D. N. 1992. Quantal analysis using maximum entropy noise deconvolution. *J. Neurosci. Methods* 44:47–57.

Kullmann, D. M. 1993. Quantal variability of excitatory transmission in the hippocampus: implications for the open probability of fast glutamate-gated channels. *Proc. R. Soc. Lond. (Biol.).* 253:107–116.

Kullmann, D. M., and R. A. Nicoll. 1992. Long-term potentiation is associated with increases in quantal content and quantal amplitude. *Nature.* 357:240–244.

Liao, D., A. Jones, and R. Malinow. 1992. Direct measurement of quantal changes underlying long-term potentiation in CA1 hippocampus. *Neuron.* 9:1089–1097.

Lilliefors, H. W. 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Statist. Assoc.* 62:399–402.

Lilliefors, H. W. 1969. On the Kolmogorov-Smirnov test for exponential distributions with mean unknown. *J. Am. Statist. Assoc.* 64:387–389.

Malinow, R. 1991. Transmission between pairs of hippocampal slice neurons: quantal levels, oscillations, and LTP. *Science.* 252:722–724.

Mason, A. L., and C. B. Bell. 1986. New Lilliefors and Srinivasan tables with applications. *Commun. Statist. Simul. Comp.* 15:451–477.

McLachlan, E. M. 1978. The statistics of transmitter release at chemical synapses. *Int. Rev. Physiol. Neurophysiol. III.* 17:49–117.

McLachlan, G. J. 1987. On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Appl. Statist.* 36:318–324.

Nikullin, M. S., and P. E. Greenwood. 1996. A Guide to Chi-Squared Testing. Wiley, New York.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. 1992. Numerical Recipes in C, 2nd Ed. Cambridge University Press, Cambridge.

Raastad, M., and R. Lipowski. 1996. Diversity of postsynaptic amplitude and failure probability of unitary excitatory synapses between CA3 and CA1 cells in the rat hippocampus. *Eur. J. Neurosci.* 8:1265–1274.

Rao, C. R. 1970. Advanced Statistical Methods in Biometric Research. Hafner Publishing Company, Darien, CT.

Smith, B. R. 1993. Mixture estimation and the neurophysiological quantal hypothesis. *Biometrics.* 49:643–651.

Smith, B. R., J. M. Wojtowicz, and H. L. Atwood. 1991. Maximum likelihood estimation of non-uniform transmitter release probabilities in neurons of the central nervous system. *J. Theor. Biol.* 150:457–472.

Srinivasan, R. 1970. An approach to testing the goodness-of-fit of incompletely specified distributions. *Biometrika.* 57:605–611.

Stephens, M. A. 1974. EDF statistics for goodness of fit and some comparisons. *J. Am. Statist. Assoc.* 69:730–737.

Stevens, C. F. 1993. Quantal release of neurotransmitter and long-term potentiation. *Cell.* 10(Suppl.):55–63.

Stratford, K. J., J. J. Jack, and A. U. Larkman. 1997. Calibration of an autocorrelation-based method for determining amplitude histogram reliability and quantal size. *J. Physiol. (Lond.).* 505:425–442.

Stricker, C., A. C. Field, and S. J. Redman. 1996. Statistical analysis of amplitude fluctuations in EPSCs evoked in rat CA1 pyramidal neurones in vitro. *J. Physiol. (Lond.).* 490:419–441.

Stricker, C., S. Redman, and D. Daley. 1994. Statistical analysis of synaptic transmission: model discrimination and confidence limits. *Biophys. J.* 67:532–547.

Stuart, A., and J. K. Ord. 1991. Kendall's Advanced Theory of Statistics, Vol. 2, 5th Ed. Oxford University Press, New York.

Torii, N., T. Tsumoto, L. Uno, A. V. Astrelin, and L. L. Voronin. 1997. Quantal analysis suggests presynaptic involvement in expression of neocortical short- and long-term depression. *Neuroscience.* 79:317–321.

von Mises, R. 1964. Mathematical Theory of Probability and Statistics. Academic Press, New York.

Wilks, S. S. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Statist.* 9:60–62.

Worden, M. K., M. Bykhovskaia, and J. T. Hackett. 1997. Facilitation at the lobster neuromuscular junction: a stimulus-dependent mobilization model. *J. Neurophysiol.* 78:417–428.

Xiang, Z., A. C. Greenwood, E. K. Kairiss, and T. H. Brown. 1994. Quantal mechanism of long-term potentiation in hippocampal mossy-fiber synapses. *J. Neurophysiol.* 71:2552–2556.