Examples of Scientific Problems and Data Analyses in Demography, Neurophysiology and Seismology

By

David R. Brillinger*

Technical Report No. 402
September 1993

*Department of Statistics, University of California, Berkeley, CA 94720,
brill@stat.berkeley.edu

Department of Statistics
University of California
Berkeley, California 94720

$2\pi \neq 1$

$2\pi \neq 1$

$2\pi \neq 1$

Examples of Scientific Problems and Data Analyses in Demography, Neurophysiology and Seismology

DAVID R. BRILLINGER *

Examples of scientific problems and data analyses are presented for the fields of demography, neurophysiology and seismology. The examples are connected by the involvement of space or time. The demographic problem is to display quantities derived from spatially aggregated data and associated measures of uncertainty. The neurophysiological problem is to infer the presence of complex pathways amongst groups on neurons given sequences of firing times. There are two seismological prolems; first to to determine isoseismals of recorded intensities following the Loma Prieta earthquake, the second to relate intensity and acceleration values measured at distinct locations. The statistical analyses are connected to each other by the application of smoothing in some form and by the provision of consquent graphical displays.

Key Words: Aggregated data; Birth rate; Coherence; Contours; Demography; Isoseismals; Locally weighted likelihood; Networks of neurons; Neurophysiology; Partial coherence; Point process; Seismic risk; Seismology; Spatial analysis; Time series.

1. INTRODUCTION

The interocular trauma - when the data hits you between the eyes.

---

* Department of Statistics, University of California, Berkeley, CA 94720, brill@stat.berkeley.edu

Display plays a central role in this paper. Indeed one of the goals of the work is clear inferences from figures. Examples are taken taken from seismology, demography and neurophysiology concerning data for which space and time are important covariates. Smoothing techniques play central roles, either as part of a locally weighted analysis or in order to obtain stable estimates. A variety of data types occur: time series, marked point processes, aggregated counts, sampled spatial fields and ordinal values.

The examples have various objectives. The objectives in the example taken from seismology are the automatic construction of isoseismal maps and the fitting of relationships between variables measured at different spatial locations. In the demography example the questions of whether there is a weekly effect in numbers of births and of how to display, via contours, spatially aggregated data are considered. How to display associated uncertainty is also an issue. In the example from neurophysiology, one is concerned with whether groups of neurons are firing together simply because of receiving a common stimulus or whether complex pathways are involved. The oft maligned technique of partial correlation analysis, when extended to point processes and the frequency domain, appears successful in studying the structure.

In each case the goal is contributing to a substantive problem. The approach followed is common sense assisted by statistical methodology and computer tools as available. Some particular tools are emphasized, some new ideas are presented and some problems arising mentioned.

2. SMOOTHING AND LOCAL WEIGHTING

A statistical tool that unites many studies is smoothing. Smoothing arises in various guises: borrowing strength, random effects, empirical Bayes, improved estimates, pooling, splines, penalized likelihood, smoothness priors. These techniques are based on different models and heuristics and can have different goals yet share the characteristics of combining parallel observations, observations that in some cases may be formally unrelated. Of one technique, Mallows and Tukey (1982) remark, "Knowing when to borrow and when not to borrow is one of the key aspects of statistical practice." In this work there will be smoothing across space, time, time lag, frequency and selected pairs of units.

Locally weighted likelihood analysis is a smoothing procedure applicable to variates of general type. It appears a pertinent estimation technique for general distributions varying in time and space. To begin consider the spatial case, with (x,y) referring to longitude and latitude. Suppose a variate Z has probability function $p(z\,|\theta)$ depending on an unknown parameter θ and possibly (x,y). Let $\psi(z\,|\theta)$ denote the score function, $\partial \log p/\partial\theta$. Now choose $\hat{\theta}$, the estimate of θ at location (x,y), to satisfy

$$\sum_i w_i(x,y)\ \psi(z_i\,|\hat{\theta}) = 0 \qquad (2.1)$$

for a weight function $w_i(x,y)$, with $w_i(x,y)$ depending on the distance of the point (x_i,y_i) from the location (x,y). Alternately look for θ maximizing

$$\sum_i w_i(x,y)\ \log p(z_i\,|\theta) \qquad (2.2)$$

This procedure gives $\hat{\theta}(x,y)$ an estimate depending on location (x,y).

Gilchrist (1967), Stone (1977), Brillinger (1977, 1990a,b and 1992a), Tibshirani and Hastie (1987), Cleveland and Devlin (1988), Hardle and

Tsybakov (1988), Staniswalis (1989), Cleveland et al (1992) are references for various stages of development of locally weighted analysis. Cleveland et al. (1992) distinguish two types of weights: neighborhood weights and robustness weights. The overall weight employed is the product. In the robustness case one needs an idea of residual.

In the example of Section 3, both smoothness and robustness will be important. In Section 4 of the paper an extension to spatially aggregated data will be indicated. In Section 5 the smoothing is elementary, over time lag, frequency or selected neuron pairs.

3. THE LOMA PRIETA EARTHQUAKE

After a sizeable earthquake, many measurements of effects are made in the surrounding region. Such data are crucial to studies of seismic risk and insurance, see Brillinger (1993). Strong motion seismometers are examined to see if they were triggered. In the case that they were, the maximum accelerations recorded will be noted. In addition reports of earthquake related effects are received, from selected observers, on a verbally described scale, the scale of modified Mercalli (MM) intensities. This scale provides 12 ordinal levels of increasing severity. For example the description of MM_{IV} reads

> Vibration like heavy truck; windows and dishes rattle; standing cars rock ,...

while the description of MM_{VIII} reads

> Damage slight in specially designed structures; considerable in ordinary substantial buildings, with partial collapse; great in

poorly built structures. Fall of chimneys, ...

The complete scale may be found in Bullen and Bolt (1985).

When such acceleration or intensity data are examined, there is found to be a general fall-off in severity of effect, with distance from the earthquake center. In the case of intensity data, isoseismal maps are prepared. The purpose of such maps is to show the pattern of ground-shaking and associated damage. The isoseismals are meant to be contours of equal intensity, bounding areas within which the predominant intensity is the same. The drawer seeks, for example, to draw a curve encircling all the MM values at a particular level. A difficulty is that the drawing of the contours has been subjective, see Reiter (1990).

Figures 1 and 2 present observations for the Loma Prieta event of 17 October 1989. The event took place near Santa Cruz, California. The epicenter of the earthquake is marked on the map by a large dot. A general description is provided in Bolt (1993). The event had magnitude 6.9, duration 10 seconds, and led to 63 deaths, 1300 buildings destroyed and 5.9 billion dollars damage. Figure 1 presents selected accelerations observed. Figure 2 provides selected intensities. (The values were selected in an attempt to minimize overstriking on the plots.) The maximum acceleration values of the event are taken from Boore et al. (1989). There are 266 observations in all. The MM intensity data analyzed are those employed in Stover et al. (1990). The largest MM intensity was IX. There were 921 observations in this case.

Intentions of the analysis to be presented include to develop: automatic displays of earthquake effects and relationships of the acceleration and intensity values.

To generate isoseismals automatically the data are processed by a form of locally weighted regression analysis. Figure 3 presents the results of the computations for the maximum accelerations. A grid size of 80 by 80 was employed. The analysis contours were constructed only for points lieing in the convex hull of the observation points. The general fall-off of acceleration with distance from the source of the earthquake is apparent with a bulge around Oakland. Figure 4 similarly presents the isoseismals derived from the MM intensities.

The model employed was

$$Z = \alpha + \beta x + \gamma y + noise$$

with Z either MM intensity or log(acceleration) and (x,y) location. A resistant fitting procedure was employed because of the substantial local variability in the data.

These smoothed surfaces have another use. The similar falloff in the Figures 3 and 4 may be taken advantage of. Acceleration data are often unavailable, particularly in the case of historical events, and there is an impetus in risk studies to develop estimates based on available MM intensities, see Brillinger (1993). A difficulty arising in constructing a conversion relationship however is that, even when recorded for the same event, the intensities and accelerations are usually measured at different places. The solution proposed here is to obtain smoothed values of both quantities and then to regress each on the predicted value of the other at the available locations. The smoothing has the further effect of reducing measurement error. A robust procedure is needed because of the occurrence of aberrant behaviour.

Figure 5 top is a scatter diagram of the observed log(acceleration) versus predicted intensity at sites where accelerations were recorded.

There is a suggestion of approximate linearity, with considerable scatter. A robust/resistant prediction line of fit has been added to the graph. The prediction relation determined was:

$$\ln \hat{A} = -7.60 + .912\, I_{MM} \qquad IV \leq I_{MM} \leq VIII$$

with a standard error of .673 . Here acceleration is measured in units of $g = 980 cm/sec^2$. In the case of intensity

$$\hat{I}_{MM} = 7.763 + .79 \ln A \qquad .02 \leq A \leq .50$$

with a standard error of .681 .

The computations leading to the contours appearing in Figures 3 and 4 were actually carried out via the function loess(), described in Cleveland and Devlin (1988), or Cleveland et al. (1992). An interesting question is how to assess and display the uncertainty of maps like those of Figure 3 and 4. The function loess() does produce standard errors which could be employed. Musmeci (1984) proposes the use of a bootstrap procedure, such as are discussed in Diaconis and Efron (1983). Further discussion of contour uncertainty display occurs in the following section.

In this section solutions to two problems have been provided: objective/automatic construction of isoseismals and predictive relationships for MM intensities and accelerations.

4. BIRTHS IN SASKATCHEWAN

In this study it is desired to investigate and display spatial and temporal structure. Problems include: Is there a weekly effect? Is there systematic dependence on location? How does one construct a contour display from aggregate data? How does one display contour uncertainty? How does one handle unmeasured covariates?

The data represent total counts of births and population for both space and time over census divisions. The data are for the Canadian province of Saskatchewan, which has 18 census divisions. The census division aggregated daily totals of births to women aged 25-29 for the two year period 1986-1987 are employed in the analysis. The year 1986 was picked because of the available census population totals for Census Day, 3 June 1986. More details concerning this data set may be found in Brillinger (1990a,b and 1992a).

It is claimed that there is a weekly effect in numbers of births caused by doctors intervening in the birth process in such a way as to reduce births over the weekend. Figure 6 is directed at this issue. The top display provides parallel boxplots of birth counts by day of the week, displayed by fortnight. There is a definite suggestion of reduced birth numbers on Saturday and Sunday. (A fortnight of data is displayed to bring out the effect more clearly taking advantage of the periodicity of the days of the week.) The bottom display of the figure provides the average number of births for each day of the week. A reduction on the weekend is again apparent. Plus and minus two standard error limits, based on Poisson variability, are indicated by the dashed lines.

Spatial analysis is considered next. Figure 7 shows the Saskatchewan census divisions and the 1986 counts of women aged 25 to 29 and their births. For example in the top census division there were 227 births to 1137 women aged 25 to 29. Figure 8 displays the annual birth rate per 1000 for each division in different ways. These rates are simply ratios of the census division counts of Figure 7. The top right and bottom left displays are examples of chloropleth or (what Tukey (1979) calls) patch maps. Because of the constancy across census divisions suggested by this

display procedure, Tukey (1979) exclaims that such maps "thereby lie, lie, lie.". He proposes some improvements. The bottom right display is a procedure developed in the present work. It is meant to bring out the information more effectively by seizing onto a plausible smoothness of relationship. The details follow.

It is often the case that contour plots provide more easily digested and read displays of spatial data. Because here the data are aggregated over census division it is not immediately obvious how to produce contours. In Brillinger (1990a,b and 1992b) it is proposed to handle the aggregation by employing locally weighted analysis with the weights

$$w_i(x,y) = \frac{1}{|R_i|}\iint_{R_i} W(x-u,y-v)dudv \qquad (4.1)$$

employing the kernel function

$$W(x,y) = (1-u^2)^2 \quad \text{for } |u| \leq 1 \qquad (4.2)$$

and equal 0 otherwise, where $u = b\sqrt{x^2+y^2}$ for some $b > 0$. Here R_i is the $i-th$ census division and $|R_i|$ its area. One can view $w_i(x,y)$ as representing the influence of census division i on a person at location (x,y), the influence resulting from items like travel, nutrition, climate, ethnicity, education, television and laws to the extent that these are functions of location. The specific formula (4.1) is motivated by an argument provided in the Appendix. For small R_i, $w_i(x,y)$ is the usual $W(x-x_i,y-y_i)$ with (x_i,y_i) a point in R_i. The weights, w_i, are evaluated via a Fourier transform, taking advantage of the convolution form in (4.1). The area $|R_i|$ is computed from the general formula for the area of a polygon.

Figure 9 shows the effect of varying the parameter b on the $w_i(.)$ for Census Division 18, the northern half of the province. The values of b,

for the three cases illustrated in Figure 9, correspond to no, a small amount, and a moderate amount of smoothing. Specifically the values were $b = 0, .05, .2$ as a fraction of of the width of the province. The value .2 is employed in the computations of the paper.

Alternate weights that might be employed are given in Tobler (1979) and Dyn and Wahba (1982). A procedure might be developed extending that of Eilers (1991). An advantage of the present approach weights is that they are local in character.

A naive model for the birth count B_{ijk}, out of N_i available women, is the binomial. Here i indexes census division, j takes on two values corresponding to weekend (Saturday, Sunday) or not weekend (Monday through Friday) and k refers to the two years 1986 or 1987. It is argued in Brillinger (1990a,b and 1992a) that a variety of pertinent explanatory variables, eg. diet, lifestyle, weather, environment, holidays, age structure, urbanicity, have gone unmeasured. This can lead to extra variation in the number of births beyond that of a binomial and is to be contrasted with the distance effects that w_i above is meant to handle. One way to proceed is to introduce a random effect, σz, and to model the birth probability via

$$logit\ \pi_j = \alpha + \beta I_j + \sigma z \tag{4.3}$$

with $j = 1, 2$ for weekday or not and I_j a corresponding indicator variable. The variate z is standard normal. This model is the binomial-logitnormal. In the present case the N_i are large and one can approximate the model by a logitnormal, i.e. by assuming the logits of the B_{ijk}/N_i are normal. This logitnormal model may be written

$$\log B_{ijk}/(N_i - B_{ijk}) = \alpha + \beta I_j + \varepsilon_{ijk} \tag{4.4}$$

with the ε_{ijk} independent normals of mean 0 and variance σ^2. For each

point (x,y) a different pair (α,β) is estimated by minimizing

$$\sum_{i,j,k} w_i(x,y)\,[\,\log[B_{ijk}/(N_i-B_{ijk})]-\alpha-\beta I_k\,]^2$$

analagous to (2.2).

Figure 8, bottom right, presented the estimate of the overall birthrate as a function of location. There is a suggestion of a trough including the main cities of the province, Regina and Saskatoon. The right panel of Figure 10 gives the estimate of the weekday effect $\hat{\beta}(x,y)$. This estimate is indeed positive across the province. There is some suggestion of spatial dependence and noting this is an advantage of the present approach. A 40 by 40 grid and contour procedure were employed in the analysis.

In the previous section the problem of, how to display contour uncertainty was mentioned. Consider the uncertainty of the annual rate estimate graphed in bottom right, Figure 8. The standard error, at a general point (x,y), may be estimated by the delta-method, see Rao (1973). The upper left panel of Figure 11 is a contour plot of that estimate basing contours on the same 40 by 40 grid. Contour levels range from 4 to 10 per 1000 births, and again suggest a valley containing Regina and Saskatoon. The dashed lines of the display are the contours for 150 and 170 births per 1000. The righthand top panel of Figure 11 indicates the shift in contour lines produced by adding and subtracting one standard error to the estimated rate function. Smooth bands appear about the estimated rate contours. The intention of these bands is clear, but thoughtful interpretation is needed. The bottom lefthand panel provides 10 simulations of the logitnormal process for the 150 and 170 contours. In these computations the individual estimates $\hat{\alpha}(x,y)$, $\hat{\beta}(x,y)$, $\hat{\sigma}(x,y)$ are first aggregated over census divisions, then independent realizations of the process are generated according to (4.4). The model is refit for each simulation, in bootstrap

style, and the 150 and 170 contours determined. The 170 contour realizations fall in an apparent band, but the 150 move about a fair amount with some quite wild curves. The final panel graphs the 150 and 170 birthrate contours, as dashed lines, and superposes the estimated standard errors at selected positions along these contours. This appears a simple way to take in the uncertainty, but one cannot read off the standard error estimates for all locations.

It would have been useful at this point to have a device capable of producing translucent displays and simple dynamic and zoom capabilities. This leads to the last suggestion provided in Figure 12. This figure is based on five east-west slices through the province along the horizontal lines indicated in the left display. Provided in the right hand display are ± 2 standard error limits about the estimated rate along the slice. One can imagine three layers of colored icing on the top of a cake with the topmost layer indicating uncertainty above the estimated birth rate and the second layer uncertainty below. This extends a traditional single variable means of indicating uncertainty for functions.

Geographers have been concerned with indicating the uncertainty of spatial quantities and of displaying spatial data generally. References include: Dutton (1977), Lanter and Giordano (1993), Muehrcke (1972), Tobler (1970), Beard and Buttenfield (1992).

This work has addressed the questions of forming a visual summary of spatially agregated birth data and of whether there were day of the week effects present. Re the former, the goal was an easily read display. Re the latter, concern was to examine a hypothesis. Figures 6 and 10 provided evidence for an effect. A sidelight of the latter work was the realization, based on the idea that there was a weekday effect in births, that

perhaps the wrong year's data had been provided initially. For the data first sent, Friday was found to have a lower birth count. On checking it turned out that 1987 had been sent instead of the desired census year, 1986.

5. ANALYSIS OF AUDITORY PATHWAYS

The next example is taken from the field of neurophysiology. Recordings of neuronal activities of the cat hearing system were made with microelectrodes inserted into the auditory thalamus. The thalamus is the major source of sensory information for the cerebral cortex. The firing times of up to 8 single units were measured. The basic phenomenon corresponds to a continuous time series, but near identical spikes stand out when records are examined. The times of these spikes are thought to represent the information transmission process. In the experiments the stimulus consisted of a burst of white noise 200ms in duration. The noise was applied every second for 300 to 400 deliveries. In collecting the data concern was with pathways amongst groups of neurons. Details on the experimental methods are given in Villa (1988,1990).

Figure 13 top provides rastor plots for one neuron in a case when it is being stimulated and also in a case when it is firing spontaneously. Such plots provide basic insight re the circumstance. The plots are prepared by stacking the sequences of firing times that occur after each application of the stimulus, successively above each other. Consider the top left display of Figure 13. One sees a vertical dark band just to the right of lag 0. This corresponds to the neuron firing rapidly for a brief period after each commencement of the noise stimulus. This is the *on response*. Just to the right of lag 200 milliseconds, one sees a faint vertical band. This corresponds to firings taking place corresponding to stopping the noise -

the *off response*. On the other hand the points in the top right display appear random. In this case no deliberate stimulus is being applied.

The second row of Figure 13 provides estimates of the crossintensity function in the stimulated and unstimulated cases. The crossintensity function, $h_{NM}(u)$, of stationary point processes M and N is defined via

$$Prob\{N \text{ point in } (u,u+du) \mid M \text{ point at } 0\} = h_{NM}(u)du \qquad (5.1)$$

for small du. This statistic may be thought of as a histogram, simply counting the number of points in successive narrow vertical strips in the rastor plot in the display above. As seen from (5.1), it provides an estimate of the rate of firing of the neuron lag u units after stimulation. The "on" and "off" peaks rising above the upper confidence level just after lags 0 and 200 respectively in the left display correspond to the stimulus affecting the neuron. In this display one also sees an indication of inhibition after the initial excitation. For convenience in indicating sampling uncertainty the square roots have been plotted here. The right display shows 4 "significant" values when 5 were to be expected.

The *coherency* is another parameter useful for discerning association of neuron firings with a stimulus. It is a stationary process analog of the traditional correlation coefficient. The coherency at frequency λ between spike trains $M = \{\sigma_j\}$ and $N = \{\tau_k\}$, $R_{MN}(\lambda)$, may be defined as follows

$$R_{MN}(\lambda) = \lim_{T\to\infty} corr\{d_M^T(\lambda), d_N^T(\lambda)\} \qquad (5.2)$$

where

$$d_M^T(\lambda) = \sum_{0<\sigma_j\le T} e^{-i\lambda\sigma_j}$$

$-\infty < \lambda < \infty$, is the empirical Fourier transform of the points (or spike times) of M in the interval from time 0 to time T. There a similar

definition for d_N^T. The *coherence* is the modulus-squared, $|R_{MN}(\lambda)|^2$. The coherence satisfies $0 \leq |R_{MN}|^2 \leq 1$, taking the value 0 if the trains are independent and a value near 1, the better d_M^T is able to (linearly) predict d_N^T (or vice versa).

In the case of the data of Figure 13 the process M is the sequence of times of commencement of application of the noise and the process N that of the firing times. Let V denote the interval between commencing the application of the stimulus. Because of the periodicity in the input, in assessing whether the stimulus has any effect, one can focus on the frequencies $2\pi\nu/V$, $\nu = 1, 2, \cdots$ in the output. Suppose $T = VL$, L an integer. The estimate of coherency is taken as

$$d_N^T(\frac{2\pi\nu}{V}) \Big/ \sqrt{\sum |d_N^T(\frac{2\pi j}{T})|^2}$$

where the sum is over $2\pi j/T$ *near* $2\pi\nu/V$. The denominator provides root mean squared sampling fluctuations in the neighborhood of $2\pi\nu/V$. The bottom displays of Figure 13 provide estimates of the coherences in the two cases and an approximate upper 95% null significance level computed as in Brillinger et al. (1976). Significance is substantial in the stimulated case, as was to be anticipated. The analyses of this figure are meant to illustrate elementary computations with point process data and the use of time lag and frequency domain smoothing.

Suppose that a stimulus, S, is applied simultaneously and repeatedly to neurons M and N. It can appear that the spike trains of the two neurons are related even though they are not physically connected. This is because of their joint influence by the stimulus. Consider the path diagram of Figure 14. Does it represent what is going on or is there perhaps some direct M-N connection? One means of study is partial

coherence analysis.

The *partial coherency* at frequency λ of point processes M and N, given a point process S, is defined to be

$$R_{NM\,|S} = \frac{R_{NM} - R_{NS}R_{SM}}{\sqrt{(1 - |R_{NS}|^2)(1 - |R_{MS}|^2)}} \tag{5.3}$$

suppressing the dependence on λ. It is a stationary process analog of the classic coefficient of partial correlation. It may be interpreted via

$$R_{NM\,|S} = \lim_{T\to\infty} corr\,\{d_N^T - \alpha d_S^T,\, d_M^T - \beta d_S^T\}$$

where α, β are the regression coefficients of d_N^T on d_S^T and of d_M^T on d_S^T respectively. The intent of the regression coefficients is to remove the (linear) effects of the empirical Fourier transform of S from the corresponding transforms of N and M. The partial coherence also satisfies $0 \le |R_{MN\,|S}|^2 \le 1$ and may be estimated by substituting coherency estimates into (5.3).

In the present case, the point process S refers to the equispaced (V =1.0 second apart) times at which the stimulus is applied. Figure 15 presents illustrative results for the neuron of Figure 14 and a second neuron observed at the same time. The results obtained are consistent with the schematic of Figure 14. In the case that two neurons are firing independently during spontaneous activity, the coherence and partial coherence are zero. This does seem to be the case of the bottom display of Figure 15. In the top displays, the firing activity of the two units is apparently associated at low frequencies during stimulation and the partial coherence analysis seemingly "removes" the apparent association. This example may be seen as validating the tool of partial coherence analysis, in the sense that partial coherence attempts to estimate the spontaneous

coherence and seems to have done so here. Various other examples are given in Brillinger et al. (1976) and Brillinger and Villa (1993).

The just described work referred to individual pairs of neurons. Consider now groups of neurons. The thalamic nuclei, belonging to the cat auditory pathways include the medial geniculate body (MGB) and the reticular nucleus of the thalamus (RE). The medial division of the MGB will be denoted M. It is of interest to assess whether relationships occur between neuron pairs, one neuron in RE and one in M. One is concerned with the existence of connections between the divisions RE and M. Figure 16 provides a pertinent schematic. To study the question an average or pooled coherence,

$$\sum_{j \varepsilon RE,\, k \varepsilon M} |R^T_{jk}(\lambda)|^2/32 \tag{5.4}$$

was computed over 32 available pairs. Figure 17 presents interesting results. During stimulation there is apparently association between the two groups of neurons, see top left panel. The bottom two panels suggest that there is not much association in the absence of stimulation. The top right panel however presents evidence for an association of the two groups of neurons remaining after the (linear) effects of the stimulus have been "removed". The discharge pattern of some pairs of neurons may have been affected selectively by the external noise stimulation. It may be speculated that these rate sensitive effects represent a correlate of basic mechanisms of learning as suggested by Hebb (1949), see Brillinger and Villa (1993).

Because the same neurons are involved in various of the coherences averaged, the approximate distribution of the average is not elementary. Hence the null level employed in this figure is the expected value of the

empirical coherence under the null distribution. In the averaging, variance stabilization of the coherence estimates was studied but had no apparent effect as the quantities were near 0.

It is notable how well a linear technique works in this highly non-linear situation. It would be worthwhile to develop a full likelihood analysis for this situation, extending that of Brillinger (1992b).

The work of this example has been directed at seeking new neurophysiological knowledge and there are hints of some.

6. GENERAL COMMENTS

The examples have in common a concern with a scientific problem, function estimation, smoothing and graphic displays. The ease of graphics for a contemporary data analyst is in a word "liberating". In this connection the productivity of the statistician/data analyst is quite astonishing. The typist, draftsman, programmer and computer operator have all gone. This increase in productivity goes hand in hand with the digitization of science. Steadily new data types are appearing in computer digestable form. In consequence the variety of fields that a statistician can contribute to appears unlimited. Further, because of telnet and email and ftp it seems that one can work with the "best" collaborator in the world. There is no need to work solely with substantive scientists on one's own campus.

There are continually hardware, software and wetware advances. But, one does wonder are not the advertisers and moviemakers better off than statisticians today in terms of graphic tools? The data analyst has to be jealous when he sees those people's dynamic, color, manipulation and multimedia capabilities.

ACKNOWLEDGEMENTS

Lindy Brewer of the United States Geological Survey provided the MM intensity data for the Loma Prieta event. Statistics Canada provided the Saskatchewan population and birth data. The spike train analyses represent collaborative work with Dr. Alessandro Villa, University of Lausanne, Switzerland. Rick Becker, Bill Cleveland, Waldo Tobler, the Editor, the Associate Editor and the Referees made helpful and encouraging comments. The computations described could not have been carried out without the help of Leo Breiman and Phil Spector of Berkeley's Statistical Computing Facility. Partial support was received from the NSF grants DMS-9208683 and DMS-9300002.

REFERENCES

Beard, M.K. and Buttenfield, B.P. (1992). "Spatial, statistical and graphical dimensions of data quality," Pp. 408-415 in Proc. 24th Symposium on the Interface (Ed. H.J. Newton). Interface Foundation of North America.

Bolt, B.A. (1993), Earthquakes, Freeman, New York.

Boore, D.M., Seekins, L. and Joyner, W.B. (1989), "Peak accelerations from the 17 October 1989 Loma Prieta earthquake," Seismological Research Letters 60, 151-166.

Bos, L.P. and Salkauskas, K. (1989), "Moving least-squares are Backus-Gilbert optimal," J. Approximation Theory. 59, 267-275.

Brillinger, D. R. (1977), "Discussion" of Stone (1977), Ann. Statist. 5, 622-623.

Brillinger, D. R. (1990a), "Spatial-temporal modelling of spatially aggregate birth data," Survey Methodology Journal 16, 255-269.

Brillinger, D.R. (1990b), "Mapping aggregate birth data," pp. 73-83 in Analysis of Data in Time (Eds. A.C. Singh and P. Whitridge). Statistics Canada, Ottawa.

Brillinger, D.R. (1992a), "Locally weighted analysis of spatially aggregate birth data: uncertainty estimation and display," Pp. 71-79 in Symposium 91, Spatial Issues in Statistics, Statistics Canada, Ottawa.

Brillinger, D.R. (1992b), "Nerve cell spike train data analysis: a progression of technique," J. Amer. Statist. Assoc. 87, 260-271.

Brillinger, D.R. (1993), "Earthquake risk and insurance," Environmetrics 3, 361-381.

Brillinger, D.R., Bryant, H.L. Jr. and Segundo, J.P. (1976), "Identification of synaptic interactions," Biological Cybernetics 22, 213-228.

Brillinger, D.R. and Villa, A.E.P. (1993), "Examples of the investigation of neural information processing by point process analysis," Advanced Methods of Physiological System Modelling, Vol. 3 (Ed. V.Z. Marmarelis). Plenum, New York.

Bullen, K.K. and Bolt, B.A. (1985), An Introduction to the Theory of Seismology, Cambridge University Press, Cambridge.

Cleveland, W. S. and Devlin, S. J. (1988), "Locally weighted regression: an approach to regression analysis by local fitting," J. Amer. Statist. Assoc. 83, 596-610.

Cleveland, W.S., Grosse, E. and Shyu, W.M. (1992), "Local regression models," Pp. 300-376 in Statistical Models in S, (Eds. J.M. Chambers and T.J. Hastie). Wadsworth, Pacific Grove.

Diaconis, P. and Efron, B. (1983), "Computer-intensive methods in statistics," Scientific American, 116-179.

Dutton, G. (1977), "An extensible approach to imagery of gridded data," Computer Graphics 11, 159-169.

Dyn, N. and Wahba, G. (1982), "On the estimation of functions of several variables from aggregated data," SIAM J. Math. Anal. 13, 134-152.

Eilers, P.H.C. (1991), "Nonparametric density estimation with grouped observations," Statistica Neerlandica 45, 255-269.

Gilchrist, W. G. (1967), "Methods of estimation involving discounting," J. Royal. Stat. Soc. 29, 355-369

Hardle, W. and Tsybakov, A.B. (1988), "Robust nonparametric regression with simultaneous scale curve estimation," Ann. Statist. 16, 120-135.

Hebb, D.O. (1949), The Organization of Behavior, Wiley, New York.

Lanter, D. and Giordano, A. (1993), "Alternative contour symbiology and their uses in bivariate mapping," Preprint Dept. of Geography, University of California, Santa Barbara.

Mallows, C.M. and Tukey, J.W.(1982), "An overview of techniques of data analysis emphasizing its exploratory aspects," Pp. 111-172 in Some Recent Advances in Statistics (Eds. J. Tiago de Oliveira and B. Epstein). Also Pp. 891-968 in Collected Works of J. W. Tukey, Vol. 4 (Ed. L.V. Jones). Wadsworth, Pacific Grove.

Muehrcke, P. (1972), "Thematic cartography," Resource Paper 19, Associates of American Geographers, 66pp.

Musmeci, F. (1984), "A method for drawing confidence bounds on seismic contour maps," Italian National Agency for Nuclear and Alternative Energy Sources (ENEA), Rome.

Reiter, L. (1990), Earthquake Hazard Analysis, Columbia University Press, New York.

Rao, C.R. (1973), Linear Statistical Inference and Its Applications, Second Edition, Wiley, New York.

Staniswalis, J. G. (1989), "The kernel estimate of a regression function in likelihood-based models," J. Amer. Statist. Assoc. 84, 276-283.

Stone, C. (1977), "Consistent nonparametric regression," Ann. Statist. 5, 595-620.

Stover, C. W., Reagor, B. G., Baldwin, F. and Brewer, L. R. (1990), "Preliminary isoseismal map for the Santa Cruz (Loma Prieta), California, earthquake of October 18, 1989 UTC," Open-File Report 90-18, National Earthquake Information Center, Denver.

Tibshirani, R. and Hastie, T.J. (1987), "Local likelihood estimation," J. Amer. Statist. Assoc. 82, 559-567.

Tobler, W. (1970), "Data display and presentation," Environmental Information Systems (Ed. R. F. Tomlinson). IGU/COGDSP Publication. University of Saskatchewan, Saskatoon.

Tobler, W. R. (1979), "Smooth pycnophylactic interpolation for geographical regions," J. Amer. Statist. Assoc. 74, 519-536.

Tukey, J.W. (1977), Exploratory Data Analysis, Addison-Wesley, Reading.

Tukey, J.W. (1979), "Statistical mapping: what should not be plotted," Pp. 18-26 in Proc. 1976 Workshop on Automated Cartography. DHEW Publication No. (PHS) 79-1254. Included in the Collected Works of J.W. Tukey, Vol. 5 (1988), (Ed. W.S. Cleveland). Wadsworth, Pacific Grove.

Villa, A.E.P. (1988), Influence de l'Ecorce Cérébrale sur l'Activité Spontanée et Evoquée du Thalamus Auditif du Chat. Unpublished thesis, University of Lausanne, Faculty of Science.

Villa, A.E.P. (1990), "Functional differentiation within the auditory part of the thalamic reticular nucleus of the cat," Brain Research Reviews 15, 25-40.

APPENDIX

A justification for the weights (4.1) follows.

Suppose that there is a random function $Z(x,y)$ with second moment function $E\{Z(x,y)\,Z(u,v)\} = m(x,y;u,v)$. Suppose that one wishes to estimate the value of Z at location (x,y) given the aggregate data values

$$z_i = \iint_{R_i} Z(u,v)dudv$$

$i = 1,...,I$, via a linear combination $\sum_i w_i(x,y)z_i$. The minimum mean squared error solution is given by

$$\Sigma \mathbf{w} = \boldsymbol{\sigma}$$

with $\mathbf{w} = [w_i(x,y)]$ and

$$\Sigma = [\int_{R_i}\int_{R_j} m(u,v;u',v')dudvdu'dv']$$

$$\boldsymbol{\sigma} = [\int_{R_i} m(x,y;u,v)dudv]$$

Suppose further that $m(x,y;u,v) = W(x-u,y-v)$ with $W(.)$ a kernel function integrating to 1. Since the R_i are disjoint, Σ is approximately $diag\{|R_i|\}$ and one has (4.1).

The kernel, $W(.)$, is seen to have an interpretation as a second-moment function which can help in its choice. When $E\{Z(.)\} = 0$, the kernel's spread corresponds to the correlation width of $Z(.)$.

A further justification might be developed in the manner that Bos and Salkauskas (1989) employ for locally weighted least squares.

Figure Legends

Figure 1. Recorded maximum accelerations from 266 available. The units are per cent of gravity, g. The solid dot indicates the epicenter of the earthquake.

Figure 2. Recorded MM intensities from 921 available.

Figure 3. Contours derived from the acceleration data. The smoothed values of the maximum acceleration at levels were obtained by the loess() function. The dots give the locations of all available measurements.

Figure 4. Contours derived from the MM intensity data. Isoseismals were obtained by loess(). The dots give the locations of the available measurements in the map region.

Figure 5. Top display graphs observed acceleration against fitted intensity value at same location. A robust line has been added. Bottom display graphs observed intensity versus fitted acceleration at same location. A robust line has been added.

Figure 6. The top display provides parallel boxplots of birth counts by day for fortnights. The bottom display plots the average numbers of births by day. The dashed lines corresponds to ±2 standard errors about the overall average.

Figure 7. Census divisions of Saskatchewan. The bottom counts provide census population of women aged 25 to 29. Upper count refers to births in 1986 to women in that age group. The census divisions with population counts of 11812 and 10372 contain the cities of Saskatoon and Regina respectively.

Figure 8. The top left panel presents birth rates per 1000 based on the numbers of Figure 7. The top right and bottom left panel are chloropleth

displays of those values. The bottom right is the estimated rate based on the model (4.3)-(4.4) and the technique described.

Figure 9. The top left panel highlights census division 18. The others panels present the function (4.1) computed for $i = 18$ for no smoothing ($b = 0$), a small amount ($b = .05$) and a moderate amount ($b = .2$) respectively.

Figure 10. The left panel is a contour plot of the estimated weekday effect, $\hat{\beta}(x,y)$ in units per 1000. The right hand panel provides an estimate of the standard error of the model, $\hat{\sigma}(x,y)$.

Figure 11. The four panels are: a) standard error of the estimated birth rate, b) ± 1 standard error contours, c) results of 10 simulations and estimations, d) the standard errors at indicated locations along the 150 and 170 levels.

Figure 12. Slices through a layer resulting from going ± 2 standard errors above and below the rate estimate. The horizontal lines of the left display indicate the locations of the slices. The central lines of the bands on the right are the estimated rates as functions of position across the slices.

Figure 13. The top displays are rastor plots for the firings of a neuron in the case of stimulation and of not. The middle is the square root of the crossintensity estimates. The dashed lines give ± 2 standard error limits. The bottom provides the coherence estimates. The dashed lines give the upper 95% null point.

Figure 14. Path diagram corresponding to stimulus S influencing neurons M and N, with no direct connections.

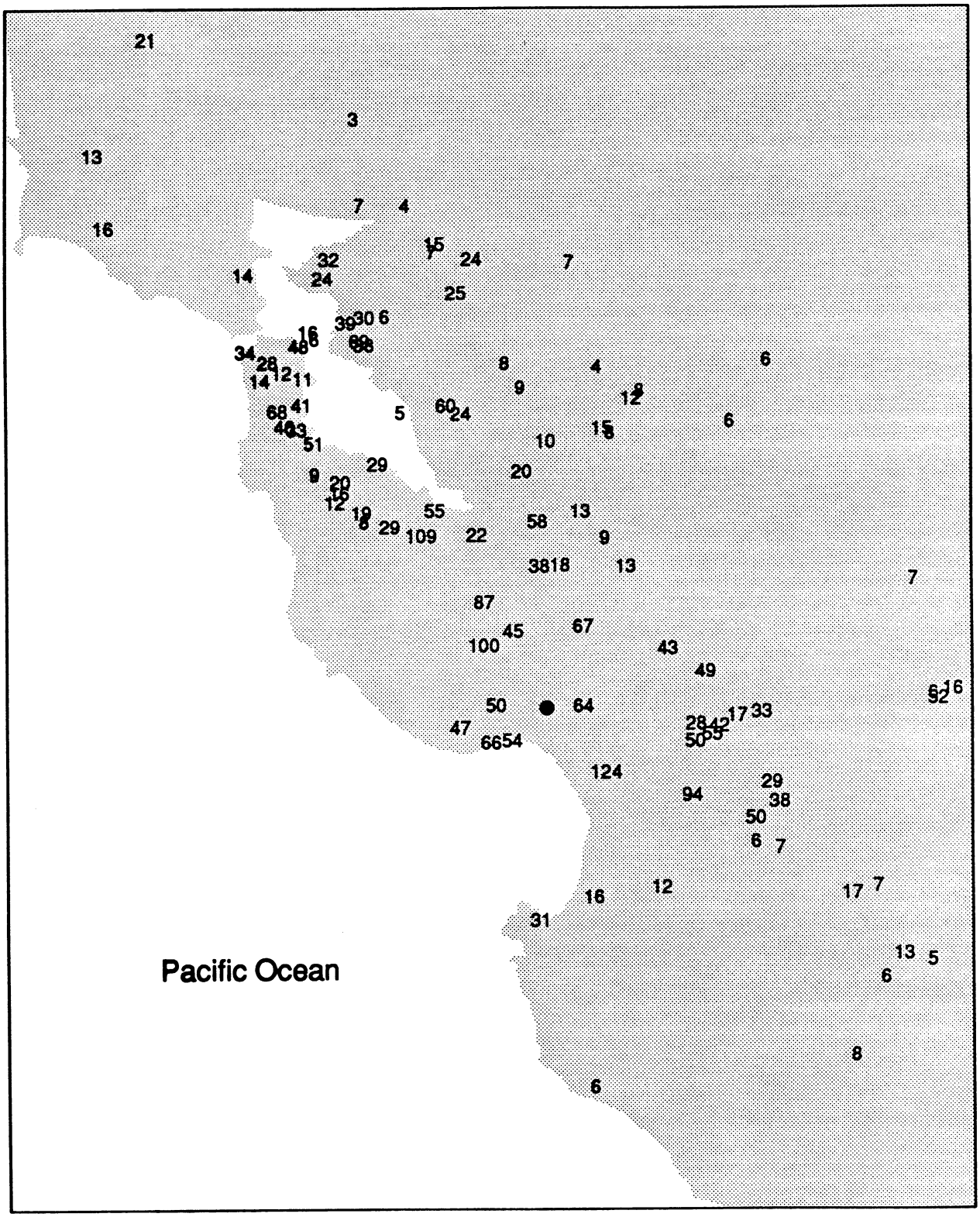
Figure 15. Left hand displays are empirical coherences of two neurons when simultaneously stimulated and when not. One of the neurons is that

of Figure 13. Right hand displays are the partial coherences in each case. Horizontal dashed lines indicate the upper 95% point of the null distribution.

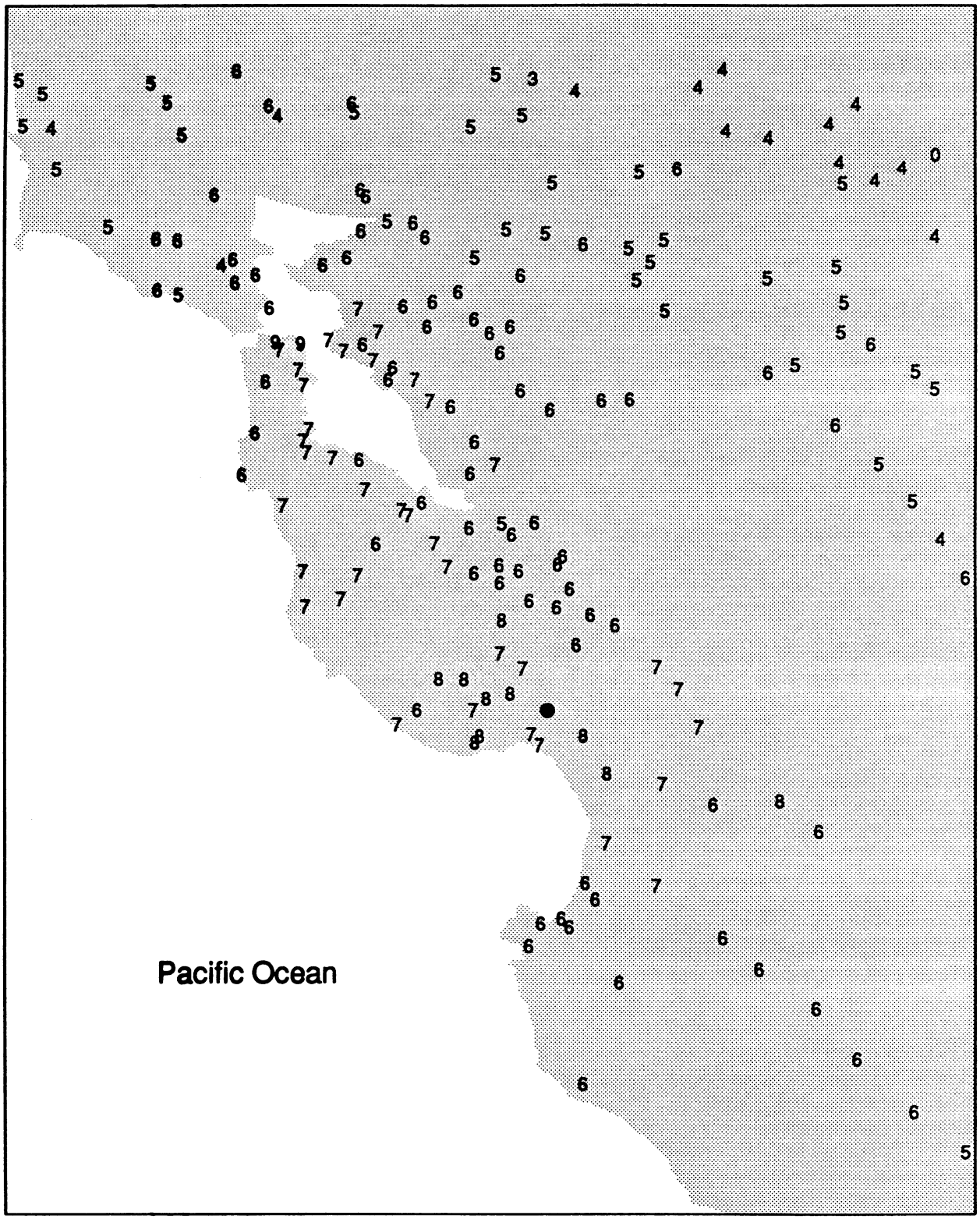
Figure 16. Path diagram suggesting stimulus S affects neurons of region RE and neurons of region M and questions whether there are connections between RE and M.

Figure 17. The average (5.4) computed first for all pairs (upper left) when stimulated, then all pairs (bottom left) when firing spontaneously. The left hand displays give similar averages, but of partial coherences. The dashed lines indicates the expected level of the average in the case of independence.

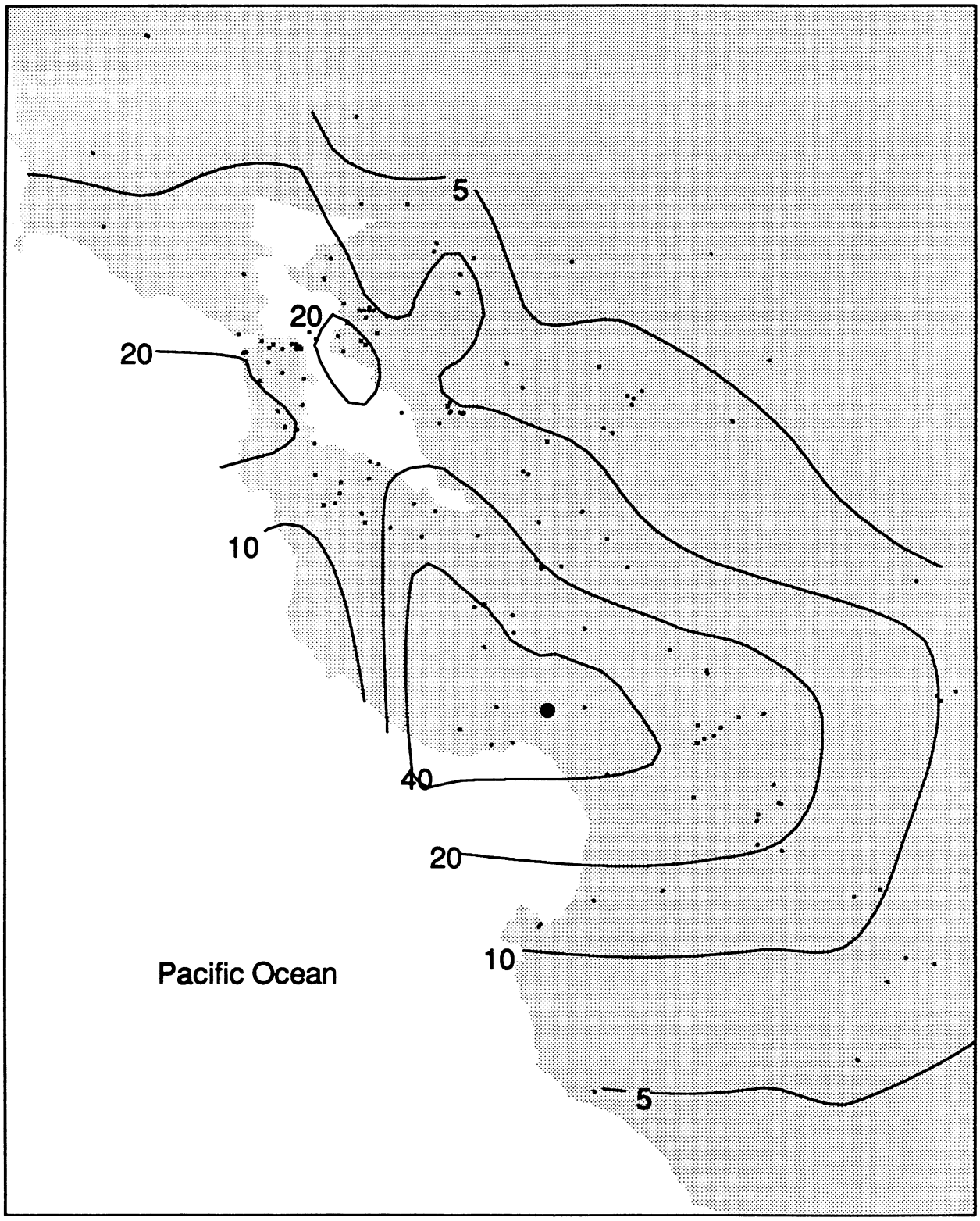
Maximum acceleration (% g)



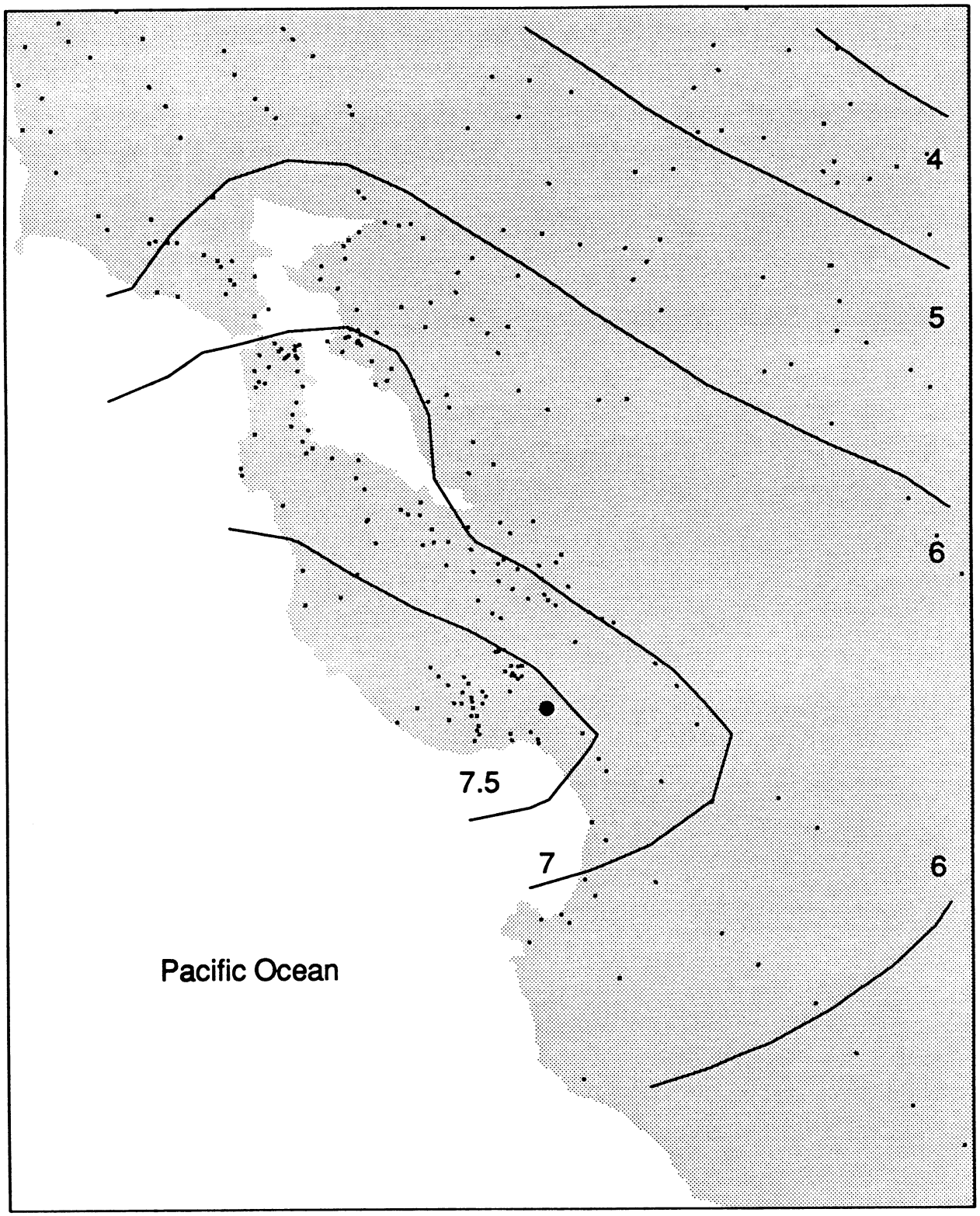
MM intensity



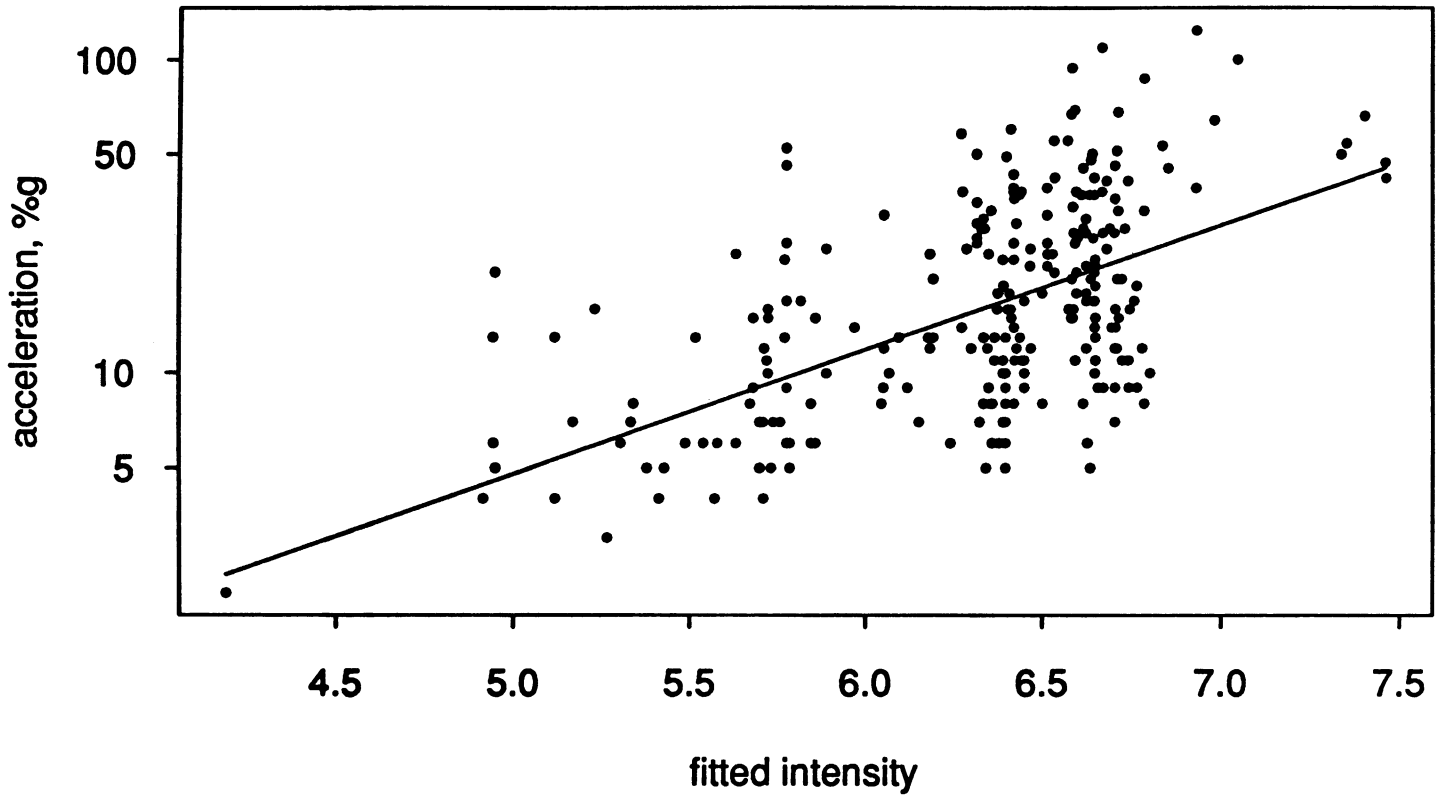
Maximum acceleration (%g)



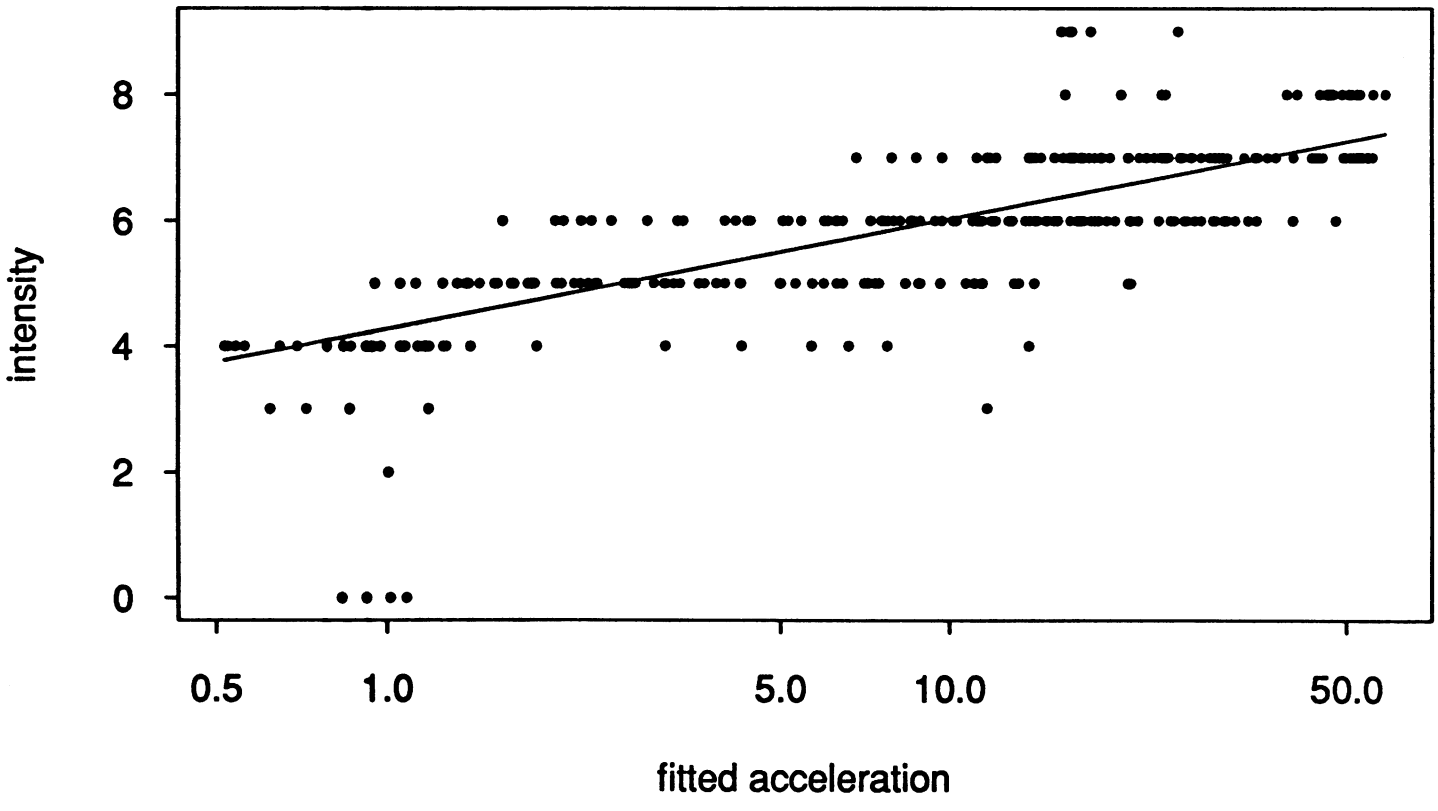
MM isoseismals



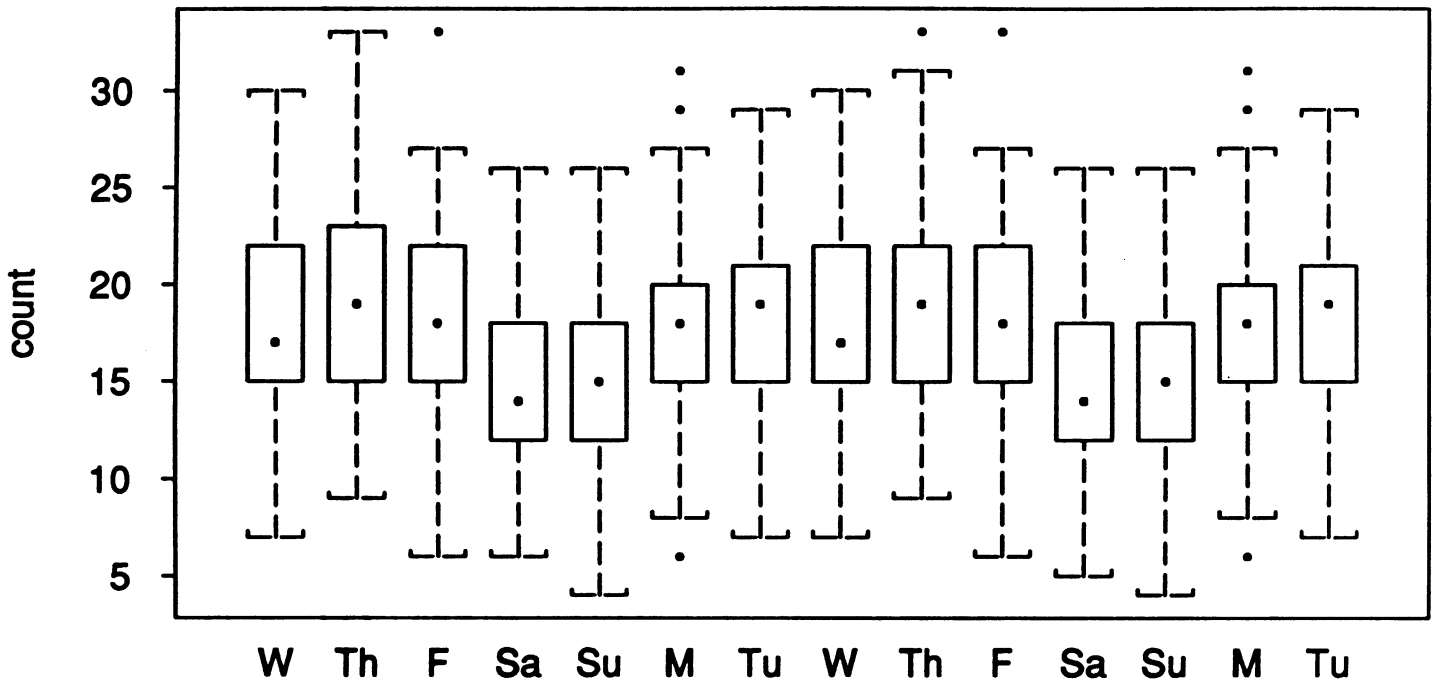
Acceleration vs. fitted MMI



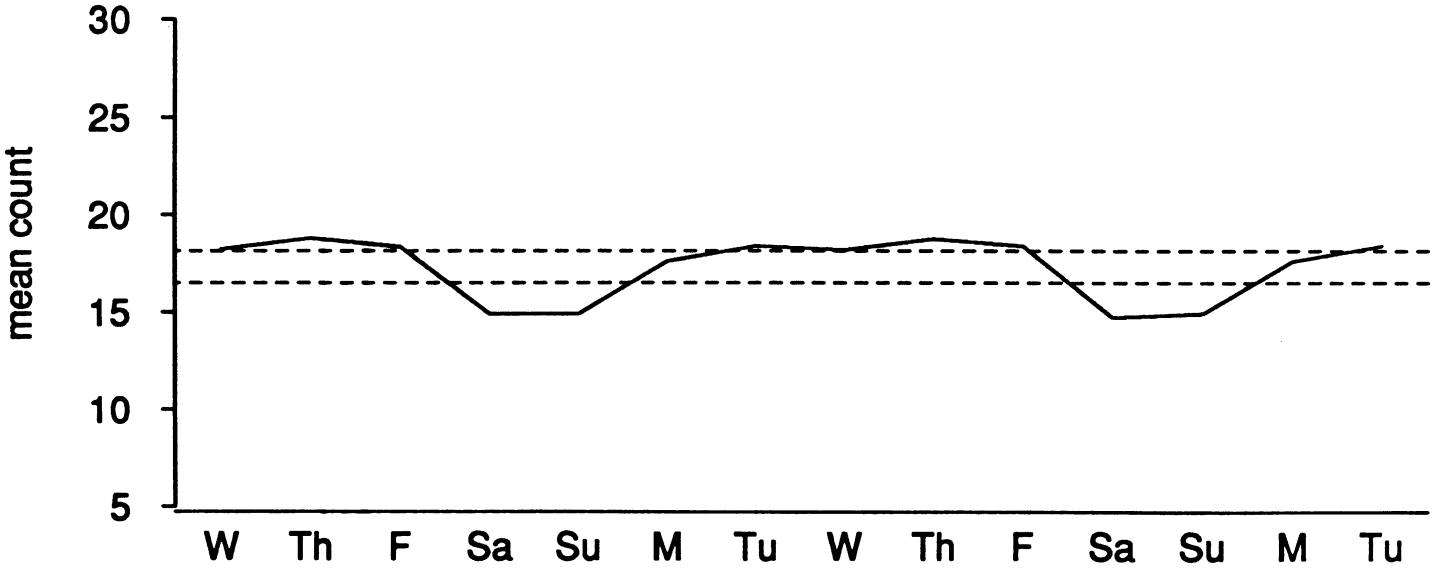
Intensity vs. fitted acceleration



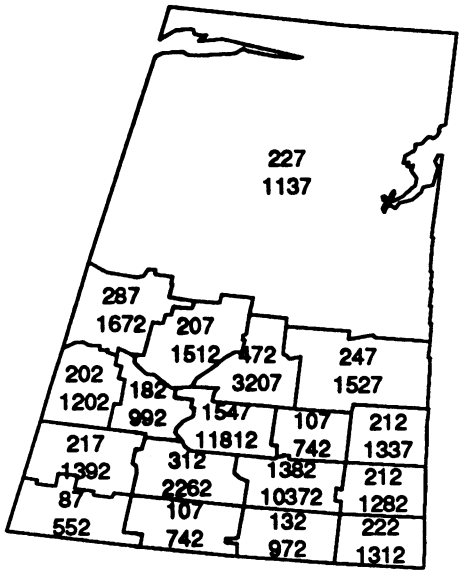
Daily births in Saskatchewan, 1986-1987

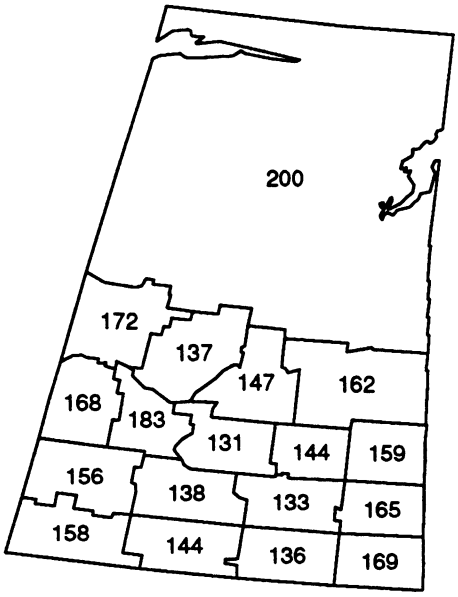


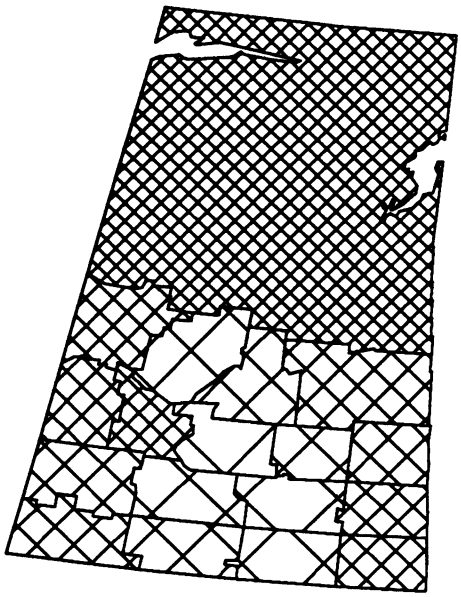
Mean daily births

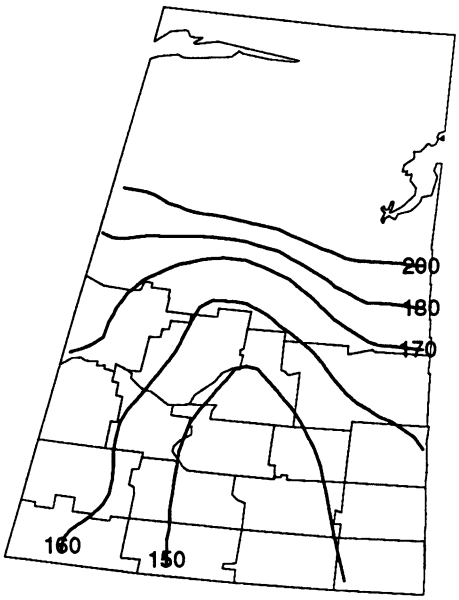


1986 births and populations, ages 25 to 29

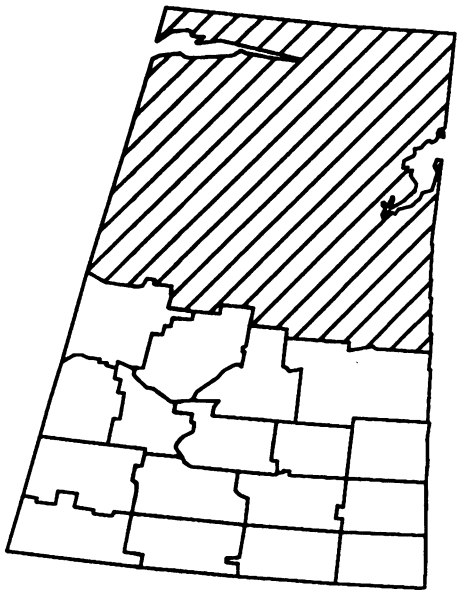


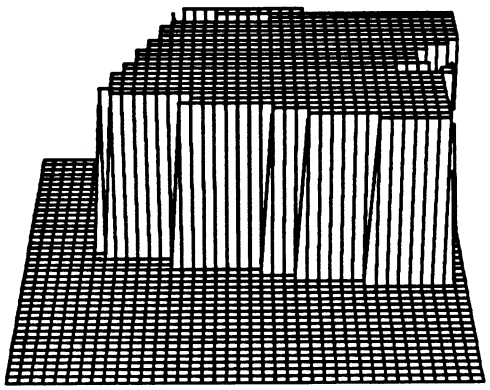

200
172
137
147
162
168
183
131
144
159
156
138
133
165
158
144
136
169

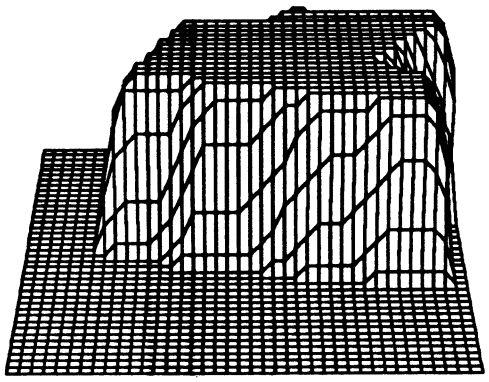


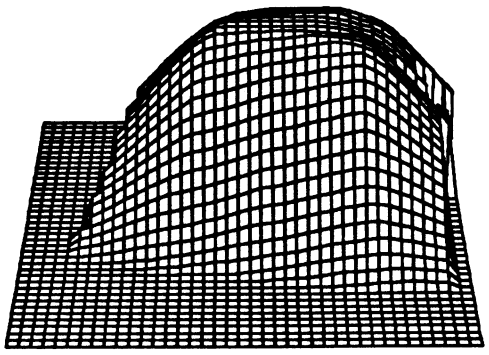

200
180
170
160
150

Census division 18

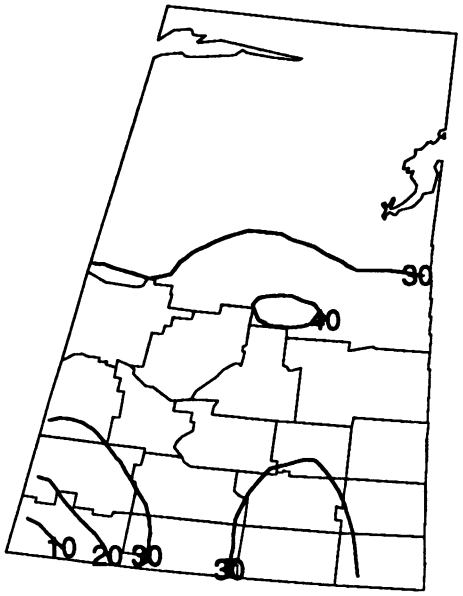




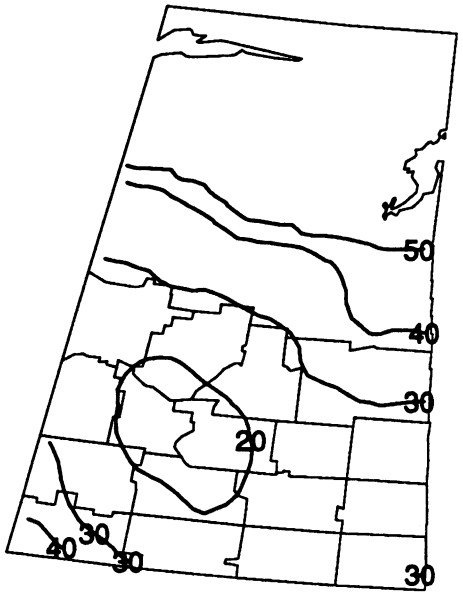




Logitnormal fit

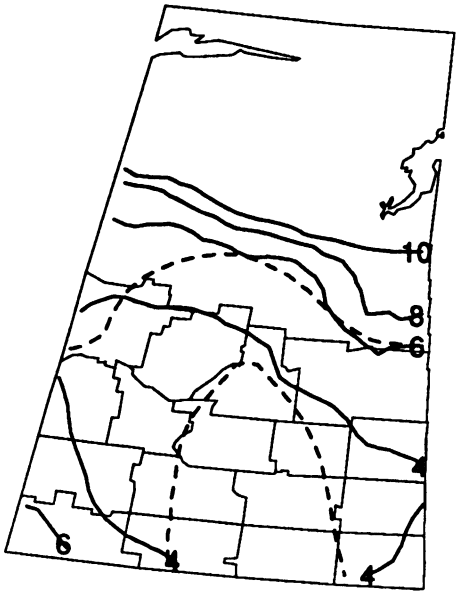


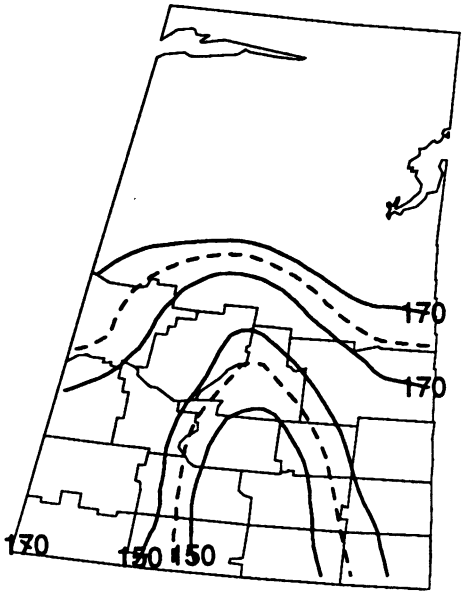
weekday effect

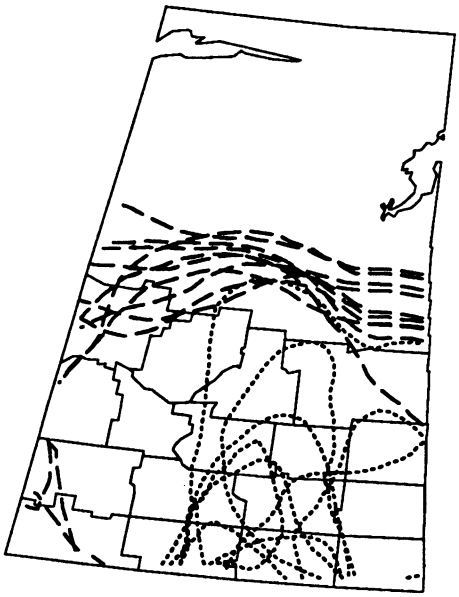


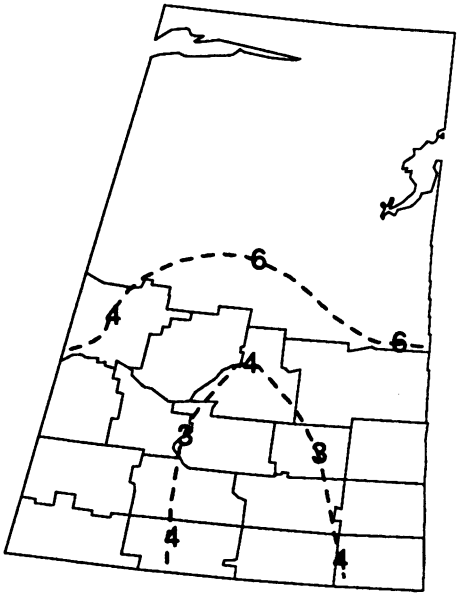
sigma

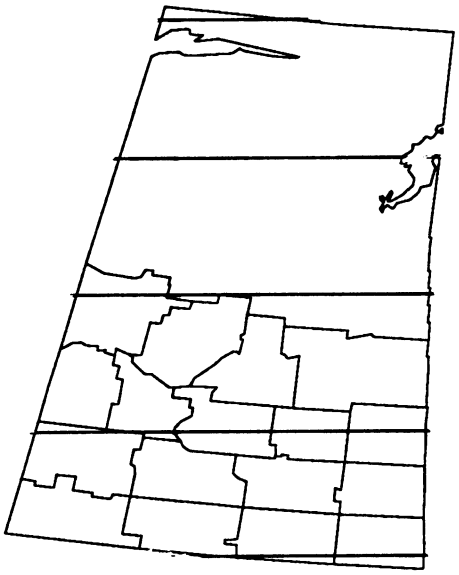
Contour uncertainty

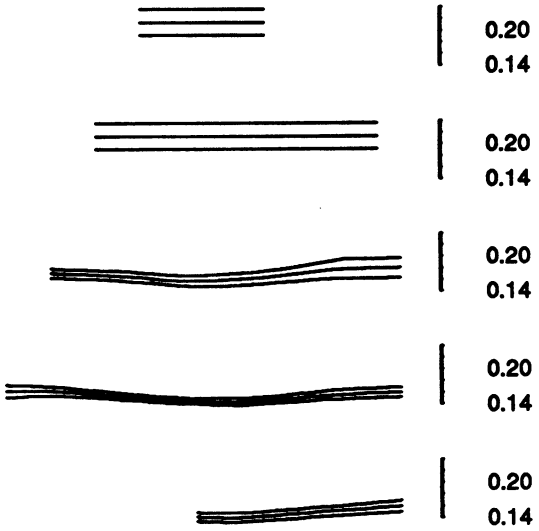


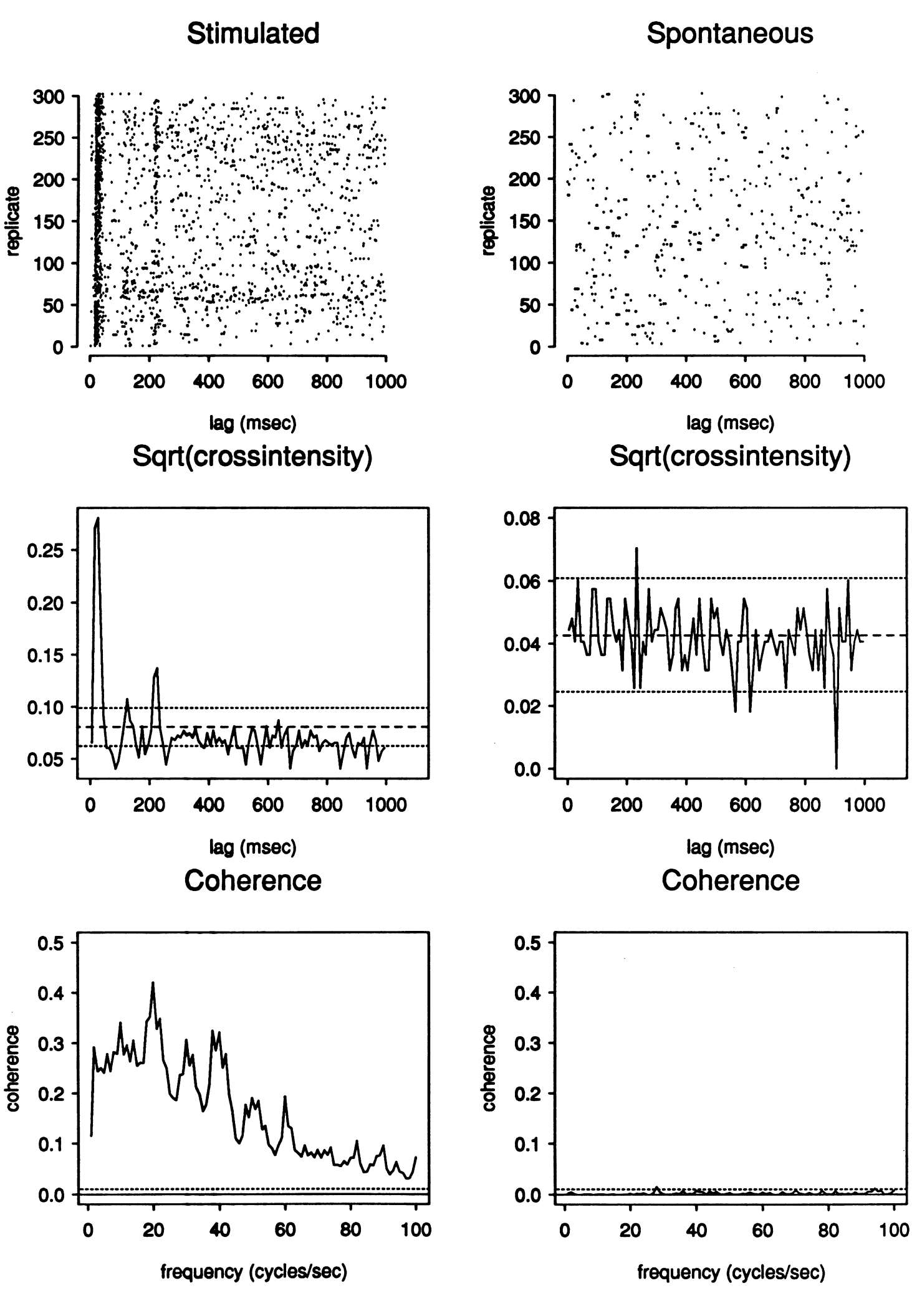


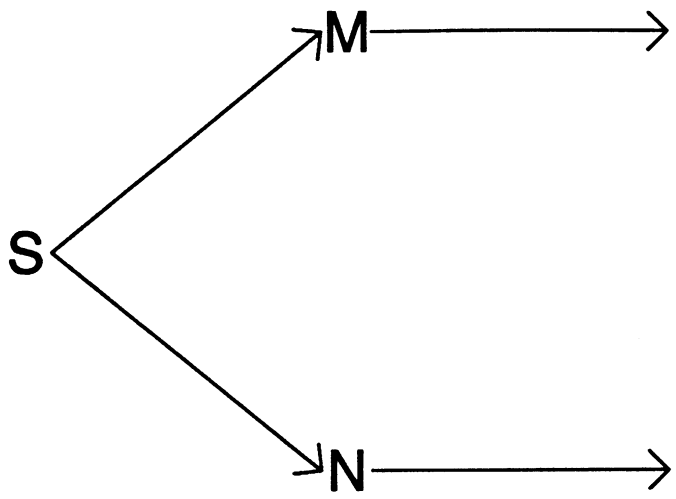


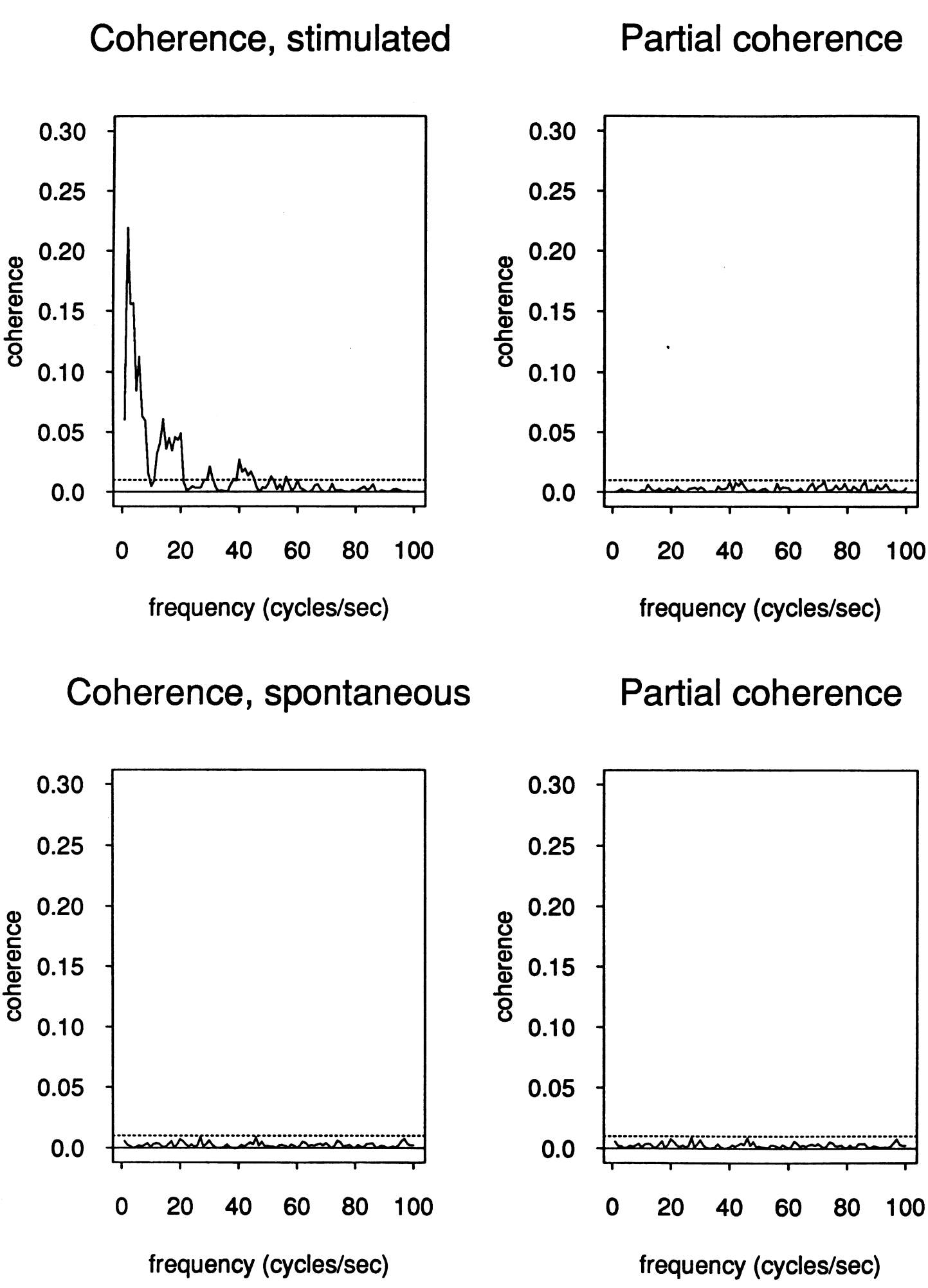


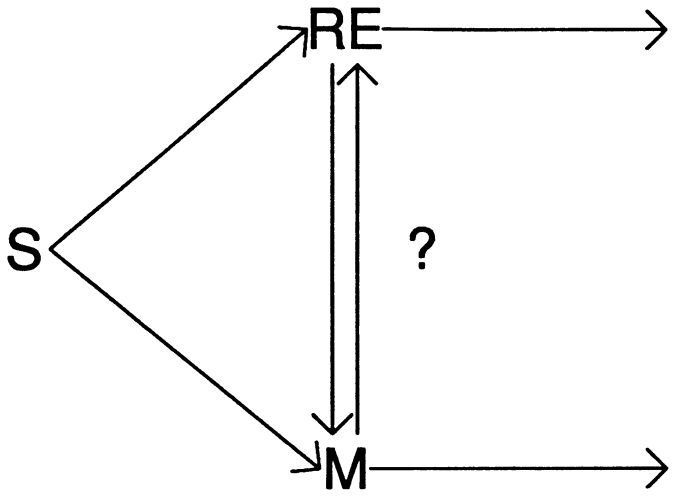


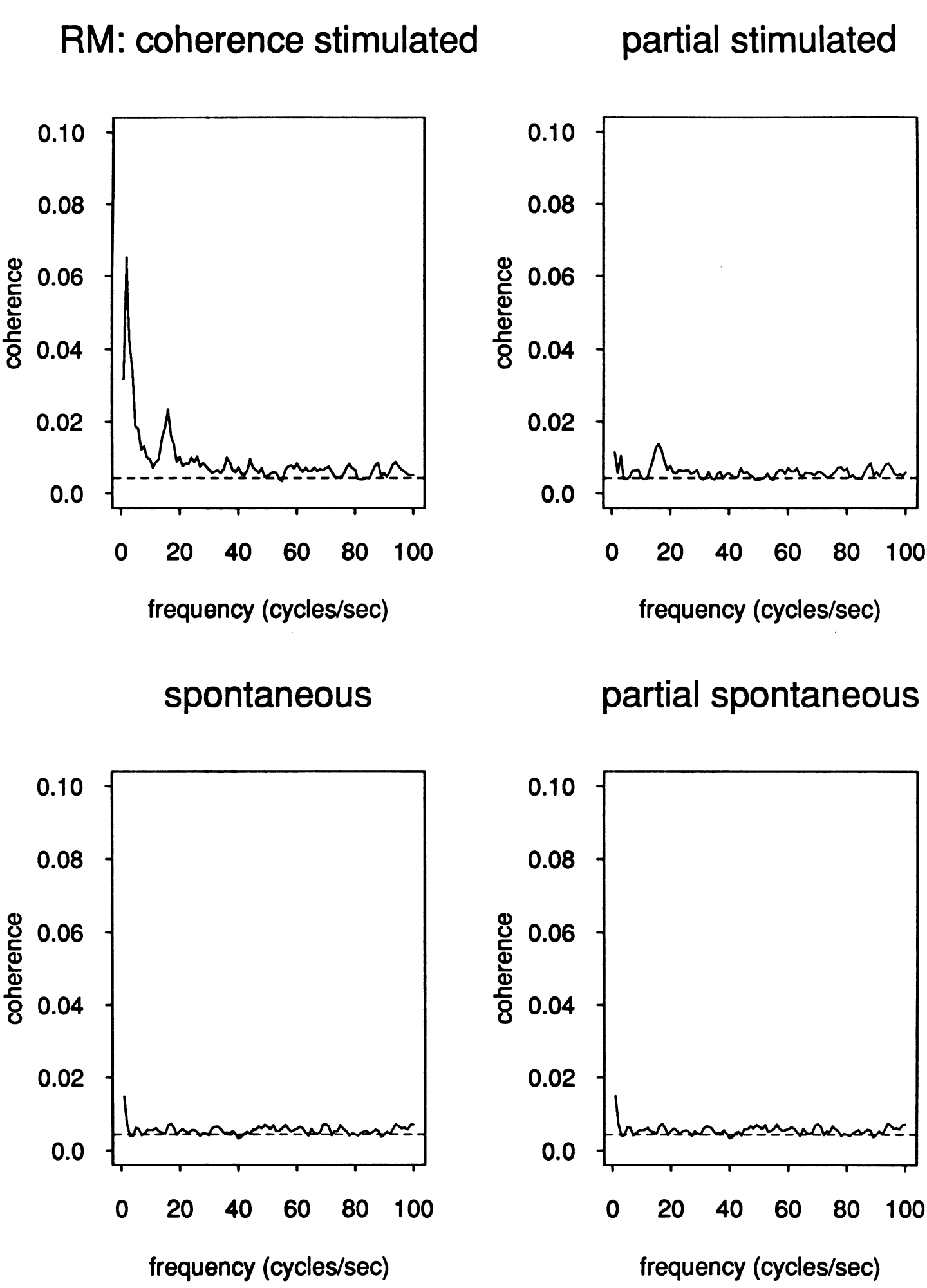

0.20
0.14
0.20
0.14
0.20
0.14
0.20
0.14
0.20
0.14


Stimulated
Spontaneous
replicate
lag (msec)
Sqrt(crossintensity)
Sqrt(crossintensity)
Coherence
Coherence
coherence
frequency (cycles/sec)


M
S
N


Coherence, stimulated
Partial coherence
Coherence, spontaneous
Partial coherence
coherence
0.30
0.25
0.20
0.15
0.10
0.05
0.0
0
20
40
60
80
100
frequency (cycles/sec)


RE
S
?
M


RM: coherence stimulated
partial stimulated
spontaneous
partial spontaneous
coherence
0.0
0.02
0.04
0.06
0.08
0.10
0
20
40
60
80
100
frequency (cycles/sec)