Statistique et analyse des données

Théo Gasser

By nonparametric methods

Statistique et analyse des données, tome 14, n° 1 (1989), p. 85-101

<http://www.numdam.org/item?id=SAD_1989__14_1_85_0>

Statistique et Analyse des Données
1989 - Vol. 14 n° 1 - p. 85 - 101

THE ANALYSIS OF BIOMEDICAL SIGNALS BY NONPARAMETRIC METHODS

Theo GASSER

Zentralinstitut für Seelische Gesundheit
D-6800 MANNHEIM

Résumé

Dans cet exposé je discute quelques difficultés rencontrées dans l'analyse de signaux biomédicaux d'un point de vue statistique (en contraste avec le point de vue ingénieur).

Une difficulté typique dans beaucoup d'applications est le choix d'un bon modèle comme base pour l'analyse de données. Souvent, les méthodes nonparamétriques offrent une alternative attractive parce qu'elles n'exigent pas un modèle choisi a priori. Une discussion détaillée est consacrée aux estimateurs à noyau pour la régression et la différentiation. En outre, une difficulté rencontrée souvent pour les signaux biomédicaux est que les résidus ne sont pas indépendants mais corrélés. Par conséquent, beaucoup de méthodes classiques de la statistique - des tests en particulier - ne peuvent être appliqués naïvement dans de telles situations.

Classification AMS : 62 G 05

Classification STMA : 04 180

Mots-clés : Régression non-paramétrique, estimateur à noyau, modélisation, analyse de signal.

Manuscrit reçu le 7. 11. 88, révisé le 14. 3 89

This work has been performed as part of the research program of the Sonderforschungsbereich 123 (project B1) at the University of Heidelberg, and was made possible by financial support from the Deutsche Forschungsgemeinschaft.

Abstract

In this paper I deal with some difficulties that we encountered when analyzing biomedical signals from a statistical point of view (as contrasted to an engineering point of view). A typical difficulty in various contexts is the choice of an appropriate model for the data analysis. Often, nonparametric methods of analysis offer an attractive alternative since they do not rely on an a priori model. Kernel estimators will be discussed in some detail for the purpose of signal extraction. A second difficulty is that residuals of biomedical signals do usually not fulfill the requirement of independence but are correlated. Therefore, many classical methods of statistics, in particular tests, cannot be applied to such signals naively.

Keywords : Nonparametric regression, kernel estimator, modeling, signal analysis.

1. Introduction

In the analysis of biomedical signals an overlap of statistical and engineering methods is found, the latter being much more frequently applied. In my view, statistics should play a more important part, firstly, by introducing more rigid arguments when comparing different approaches and, secondly, by better accounting for interindividual variability. In this paper I will focus on the following two areas.

1

The choice of model is a crucial step in a regression and in a time series situation. In particular in a biomedical context, this step may pose serious problems. Nonparametric methods can in these circumstances become an attractive alternative.

2

The analysis of biomedical signals can usually be formalized as a regression problem with correlated residuals with unknown covariance structure. The coloured noise prohibits the naive use of classical methods of inference. This has often been overlooked in the engineering literature, e.g. by applying χ^2 or Kolmogorov-Smirnov-tests as goodness-of-fit tests to correlated data.

The first problem will be illustrated and discussed in the regression and the time series context. The difficulties of parametric modeling in regression can be elucidated in a classical application such as analysis of human growth (Gasser et al., 1984), where a nonparametric technique via kernel estimation describes the data better (Gasser et al., 1985a). Therefore, recent progress on kernel estimation is reviewed, including the data-driven choice of the smoothing parameter and an integration of various well-known smoothers under the heading of design-adaptive kernel estimators. The automatic choice of the smoothing parameter shows the importance of the second problem since all methods proposed rely on independent residuals. The importance and difficulty of the model choice is also illustrated for electroencephalogram (EEG) data, where familiar parametric models (such as autoregression) failed to adequately describe the EEG of a sample of subjects (Steinberg et al., 1985).

The second problem (already discussed in the context of the choice of smoothing parameter) arises in many different applications. I will discuss one, i.e. the analysis of trial-to-trial variability of brain signals evoked by repeated stimuli. Again, the application of standard tests to test some psychophysiological hypotheses is not feasible, due to the correlation of residuals. However, the prior application of a prewhitening filter opens access to rigorous tests (Möcks et al., 1984), and this procedure can be justified in a maximum-likelihood framework (Pham Dinh Tuan et al., 1987).

2. Regression Analysis

Let us assume that data $X_i = X(t_i)$ has been recorded which can be modelled as a signal s corrupted by additive noise ε : (with $E(\varepsilon_i) = 0$)

$$X_i = s(t_i) + \varepsilon_i \qquad (i = 1, \ldots, n)$$

The temporary assumption of independent, identically distributed residuals (with variance σ^2) makes arguments simpler. The more difficult case of couloured noise will be discussed at the end of this section. When s(t) belongs to some parametric family of functions $F(t ; a_1,\ldots,a_k)$ ($a_1,\ldots,a_k$ = regression parameters) and when this parametric family is a priori known, parametric regression is the method of choice. However, even a minor misspecification of the parametric model may prevent us from describing

adequately the structure of the regression data. This is illustrated for human height growth in figure.1, where the same boy has been measured from birth to adulthood. Besides a well-known and accentuated pubertal spurt (velocity peak at about age 14), a further spurt at about age 7 can be clearly seen in the data for this child. The nonparametric fit quantifies this phenomenon nicely, whereas it is absent in the parametric fit since it is not part of the model (Preece-Baines model, the best fitting so far for growth data). However, this lack of structure leads also to bias in the pubertal period which can be seen in this child, and which has proved to be systematic (Gasser et al., 1984).

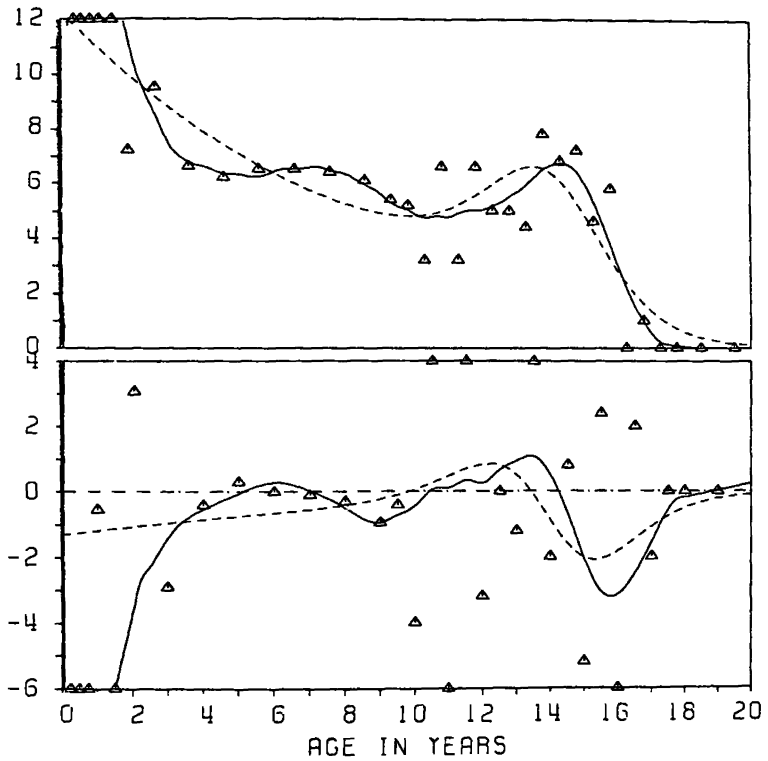

Figure 1.

Velocity (above) and acceleration (below) for height growth of a boy from birth to twenty years. Divided differences from data (Δ), kernel estimators (solid line), parametric fitting of Preece-Baines model (dotted line).

This altogether has spurred an interest in nonparametric regression methods, such as kernel estimators. As in the case of growth data, it is often of interest to estimate derivatives $s^{(v)}$ of s as well. The definition of the estimator $\hat{s}_v$ follows Gasser an Müller, 1984 :

$$\hat{s}_v(t) = \sum_{i=1}^{n} g_i X(t_i) \qquad v = 0,1,2$$

$$\text{where : } g_i = \frac{1}{(b(t))^{v+1}} \int_{s_{i-1}}^{s_i} W_v\left(\frac{t-u}{b(t)}\right) du$$

$$s_i = (t_i + t_{i+1})/2$$

The kernel function W_v and the bandwidth or smoothing parameter b(t) need to be specified. Recommendations for this choice will be given below. First, I want to discuss why to decide for kernel estimators, and not for some other nonparametric technique, such as cubic smoothing splines or k-nearest neighbour estimators. For a constant bandwidth b(t) = b one obtains the ordinary fixed-width kernel estimator. Assuming an equidistant regression design most nonparametric regression estimators can also be represented as a fixed width kernel estimator. For design points $\{t_1,...,t_n\}$, distributed according to some density f(t), a wide class of non parametric estimators can be represented as kernel estimators when defining a systematically varying bandwidth as $b(t) = b_0 . f(t)^{-\alpha}$ (these arguments are asysmptotic). This incorporates in particular ordinary kernel estimators ($\alpha = 0$), cubic, smoothing splines ($\alpha = 1/4$; Silverman, 1984) and k-nearest-neighbour estimators ($\alpha = 1$). Such a representation also enables a comparison of the merits of the different methods in terms of mean square (MSE) : it turns out that none is uniformly optimal, whereas the ordinary kernel estimator ($\alpha = 0$) is minimax optimal. A more detailed exposition of this topic may be found in Jennen-Steinmetz and Gasser (1988). For a more general discussion of nonparametric curve estimation, the reader may consult the book by Bosq and Lecoutre (1987).

For ease of presentation (and in view of theoretical results), I will concentrate on fixed bandwidth b(t) = b. A kernel of order k for estimating a derivative of order v ($v = 0, 1, 2,...$) has to satisfy the following moment conditions with $k = v + j$ ($j = 2, 4,...$) :

$$\int_{-1}^{1} W_\nu(x)\, x^m b(x) \quad \begin{cases} = 0 & m = 0,\ldots,\nu-1,\nu+1,\ldots,k-1 \\ = (-1)^\nu\, \nu! & m = \nu \\ = \beta_k \neq 0 & m = k \end{cases}$$

Asymptotic expressions for bias and variance are as follows :

$$\text{Bias}\,(\hat{s}_\nu(t)) \sim \frac{1}{k!}\, b^{k-\nu}\, s^{(k)}(t) \int W_\nu(x)\, x^k\, dx$$

$$\text{Var}\,(\hat{s}_\nu(t)) \sim \frac{\sigma^2}{nb^{2\nu+1}} \int W_\nu^2(x)\, dx$$

These formulae lead to the asymptotic integrated mean square error (IMSE), which in turn gives the rate of convergence as $O(n^{-2(k-\nu)/2k})$. The IMSE can be interpreted as a functional of W_ν. When minimizing it with respect to W_ν, optimal kernels can be derived (see figure 2 and Gasser et al., 1985b). Optimal kernels of order k are polynomials of order k for which a neat representation could be derived in terms of Legendre polynomials. Appropriate adjustments need to be made for the boundary.

A further - and crucial - parameter which needs to be specified is the bandwidth b. The optimal choice depends on the signal itself and on the variance of ε and is thus inaccessible. A subjective choice, however, has various drawbacks. This has led to a number of proposals for determining an estimate of the optimal bandwidth in a data-adaptive way (see e.g. the paper by Härdle et al., 1988 and their references). The most popular one is cross-validation, in which an estimator of IMSE is minimized with respect to b, based on one-leave-out residuals. We have worked out a different method which turned out to be flexible, efficient and fast (Gasser et al., 1988). It consists of estimating the asymptotically optimal bandwidth

$$b_{asy} = \left(\frac{1}{n} \frac{\sigma^2}{\int_0^1 s^{(k)}(t)^2 dt} \right)^{\frac{1}{(2k+1)}}$$

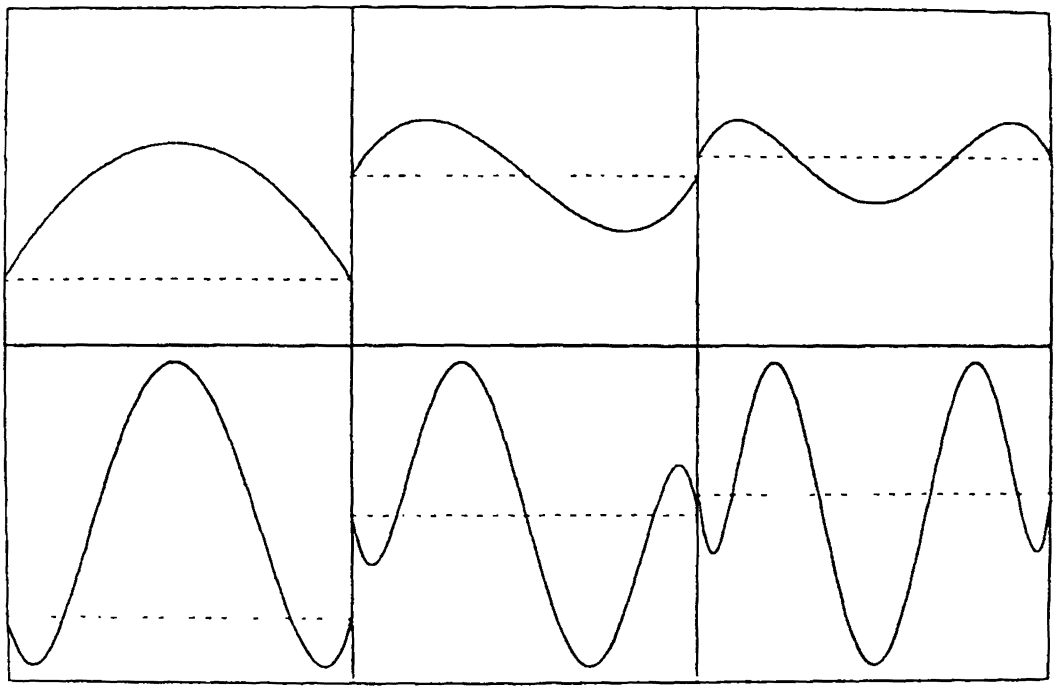

Figure 2.

Kernel of order (ν + 2) [above] and of order (ν + 4) [below]; ν = 0, 1, 2 from left to right ; dotted line at zero.

firstly by estimating σ^2 (following Gasser et al. 1986 a) and secondly by estimating $\int_0^1 s^{(k)}(t)^2 dt$ in an iterative algorithm.

Figure 3 shows in a simulated example (data = sine-wave plus Gaussian noise) that the method works well down to small sample size (here n = 25). Depicted are the kernel estimates when using the true optimal banwidth and the estimated bandwidth (the latter is typical since it had median mean square error in 401 simulated replicates).

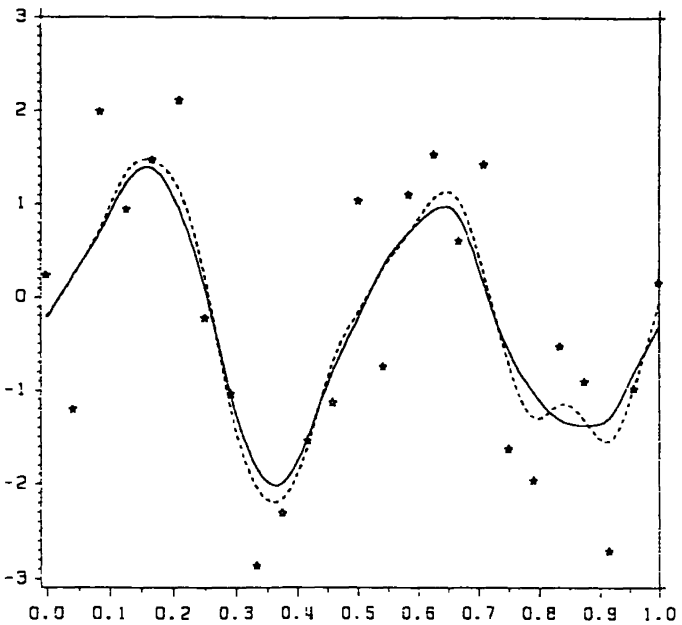

Figure 3.
Simulated data (*) with kernel estimate using truly optimal bandwidth (dotted line) or estimated optimal bandwidth (solid line).

What happens when the assumption of i.i.d. residuals is not fulfilled ? The violation of an invariant distribution (in particular of a constant variance) does not matter so much, but lack of independence does, and many biomedical signals do have coloured noise. Practically important is, that none of the methods for estimating the optimal bandwidth works for correlated errors (Hart, 1987). Regarding theory, the asymptotic expression for the bias of the kernel estimator remains unchanged, but the variance is changing (Gasser et al., 1986 b). In fact, the estimator is no longer consistent (Hart and Wehrly, 1986). Despite of these problems, kernel estimators proved to be a valuable tool for analyzing biomedical signals with correlated noise.

Figure 4 provides an example of a smoothed pH-curve (24 hour intragastric recording), leading to about 17 000 points, where parametric modeling is neither possible nor of interest. On the other hand kernel estimators provide the information of interest to gastroenterologists : a quantification of the pH-effect of meals and drugs (via derivatives) and an assessment of the pattern of pH-changes.

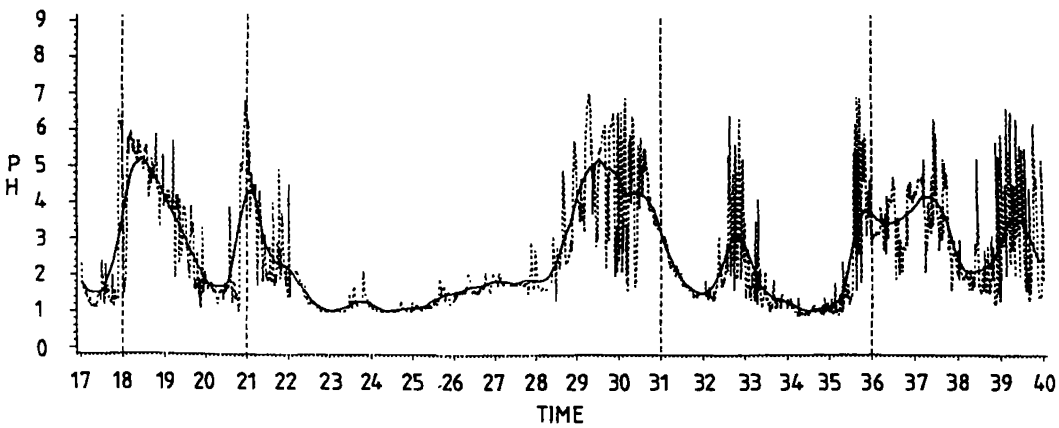

Figure 4.

Intragastric 24 h pH-recording of a healthy subject (dotted line) and kernel estimate (solid line). Vertical lines indicate the time of meals.

3. Electroencephalogram Time Series

The EEG records electrical brain activity from various (typically 4 to 20) locations on the head. Since the sampling rate is about 100 Hz, one minute of recording leads up to 120 000 data points. Any method to be considered has thus to be computationally fast. The basic assumption which is usually made is that the EEG can be modelled as a stationary vector process. Empirical studies show that stationarity holds to a good approximation for periods up to 40 seconds. The process is in most cases not too far from Gaussian so that the bulk of information about the process is contained in the spectrum / cross-spectrum matrix.

As in regression, a parametric model-oriented approach of time series analysis is in many respects attractive, and it has, therefore, attracted a lot of attention. Autoregressive (AR) modeling consists in fitting a linear p-th order scheme to the data, driven by white noise ε_t :

$$\sum_{k=0}^{p} a_k X_{t\,k} = \varepsilon_t$$

If driven by a moving average of ε_t , an autoregressive-moving average process would result. It should be noted that the parameters $(\alpha_k)_k$ themselves are difficult to interprete in an EEG context ; however, Zetterberg (1969) has proposed a transformation to determine more meaningful spectral parameters. The order p plays a role similar to bandwidth in regression, and different rules have been suggested to determine the order from the data. In a comparison for EEG series, consistent rules (Hannan and Quinn, 1979) proved to be superior for data analysis (Steinberg et al., 1985) compared to popular inconsistent rules (Akaike, 1969). The introduction of AR-fitting to selective frequency bands brought considerable advantages. However, despite these refinements, empirical results based on AR-modeling remained somewhat disappointing compared to a nonparametric approach. These empirical results refer to a comparison with broad band power (see below) in a study of retest reliability for repeated recordings, in a further study of developmental change in school-age children and a third one, trying to associate IQ with EEG parameters. In all three studies, broad band power brought sharper results compared to transformed AR-parameters (based on selective frequency range and with choice of order optimized). Nonparametric spectrum estimation consists in smoothing the periodogram, essentially also kernel estimation (Brillinger, 1975). Figure 5 shows an extreme case of model bias : 6 x 20 seconds consecutive EEG data have been analyzed by AR-modeling (using Hannan's criterion) and by nonparametric spectrum estimation. The clear two-peak-structure does not come out at all in the parametric spectrum estimates. Increasing the order of the AR-fit much beyond the optimal order would lead to a smaller bias. However, it would then be difficult to deal adequately with the large number of parameters (large compared to other subjects). In fact, such an estimate would be closer to the rationale of nonparametric fitting than parametric fitting. The traditional way of extracting parameters from the nonparametric spectrum estimate consists in summing power in prespecified frequency bands (2-6 Hz wide), selected according to neurophysiological experience. This procedure is not very appealing from a statistical point of view, but in our experience it was more advantageous than a refined AR-fitting from EEG analysis.

Most of the parametric models used in biomedical signal analysis are purely descriptive, applied for the purpose of data fitting and subsequent data reduction for further statistical analysis (AR fitting is an example).The critical remarks made with respect to a model-oriented data analysis refer only to this context and not to the one, where intrinsic biomedical knowledge is the basis for analysis ("biological modeling"). A basic problem specific for biomedical signals is that the model has to be appropriate for a sample of subjects, and often for different groups (e.g. controls and patients). It is not sufficient that a model shows a good performance for a majority of normal subjects, which may often be the case for some reasonable model. These problems turned out to be even more accentuated for EEG-analysis compared to growth curve analysis, since even normal subjects show a large variety of patterns of brain activity.

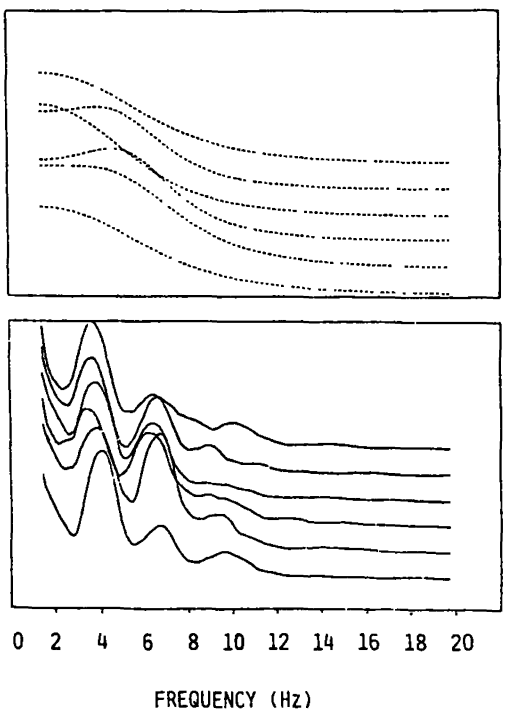

Figure 5.

Spectrum estimation of 6 x 20 seconds. of EEG data : parametric AR-fitting (above) and kernel estimation (below).

analysis ("biological modeling"). A basic problem specific for biomedical signals is that the model has to be appropriate for a sample of subjects, and often for different groups (e.g. controls and patients). It is not sufficient that a model shows a good performance for a majority of normal subjects, which may often be the case for some reasonable model. These problems turned out to be even more accentuated for EEG-analysis compared to growth curve analysis, since even normal subjects show a large variety of patterns of brain activity.

4. Analysis of Brain Potentials

The response of the brain to a stimulus is of interest in a number of biomedical fields, and this problem has also attracted the interest of statiticians (see the recent Wald lecture by Brillinger, 1988). Due to the low signal-to-noise ratio, the response to a single stimulus is not even visible, and stimuli are repeatedly given (typically from 20 to 400 times) to enhance the signal-to-noise ratio by averaging brain activity time-locked to stimuli (since the noise has random phase, it tends to cancel). The basic model for one subject, at one brain location, is the following :

$$X_j(t_i) = r(t_i) + \varepsilon_{ij}$$

$i = 1,\ldots,T$ (indexing post-stimulus time)

$j = 1,\ldots,n$ (indexing stimulus number)

X = data, r = response, ε = noise

A crucial but shaky assumption is the invariance of the response across stimuli this also justifies the use of the average $\hat{r}(t_i) = \frac{1}{n}\sum_{j=1}^{n} X_j(t_i)$.

It would be of interest to check the validity of this assumption which may hold better for some experimental paradigm compared to another. We studied the following alternative models, allowing for a stimulus-dependent response $r_j(t_i)$:

A.	$r_j(t_i) = a_j\, r(t_i)$	model with amplitude variation
B.	$r_j(t_i) \equiv r_{j-1}(t_i)$	model with slow variation
C.	$r_j(t_i) = r(t_i - \tau_j)$	model with latency variation (τ=latency)

Model C is the only one considered previously by Woody (1967) who suggested a heuristic procedure for estimating the τ_j and then to obtain a latency-corrected averagee response.

This is not the place to review our work on testing and estimation in these models (see Möcks et al., 1984 and Pham Dinh Tuan et al., 1987 for the statistical part). Rather, I will give a short outline of problems related to the colour of residuals ε which already posed problems when determining an optimal banwidth for kernel estimators. The noise ε (mainly spontaneous brain activity) is usually far from white and shares its spectral domain with that of the response r. The ideas for testing alternatives A and B are rather intuitive (Möcks et al. 1984). Regarding A, let us note that $c_j = \mathrm{cov}(X_j,r)$ is in the mean proportional to a_j. The variability of the c_j is thus a measure for the extent of amplitude variability. To obtain a test statistic for alternative A, it has to be scaled by the variability of c_i to be expected under the standard model. The main obstacle for obtaining distributional properties of the ensuing statistic is the lack of independence in the data. This is also true for a statistic derived for testing model B. This statistic is based on the plausible idea to compare the power of successive differences $(X_{j+1}(t)-X_j(t))$ with another estimator of power, based on the basic model. The tool to circumvent the problem of correlated data was in both cases the introduction of a prewhitening filter : the data are filtered with a transfer function with zero phase and with gain $[f(\nu)]^{-1/2}$, where f is an estimate for the noise spectrum. With this preliminary step, the asymptotic distribution of the above mentioned test statistics could be derived and they proved to be a good finite sample approximation in simultations.

When studying model C, the prewhitening step got firmer mathematical support (Pham Dinh Tuan et al., 1987) : based on a maximum-likelihood argument estimators for

the latencies τ_j also needed a prior prewhitening step. This proves, that the prewhitening technique is more than an intuitive trick when dealing with correlated data. Estimation of latencies is the one problem where from the engineering side Woody (1967) had proposed an estimator based on heuristic reasoning. This estimator was compared with the maximum-likelihood approach. Figure 6 shows that the latter method is superior in MSE by order of magnitude, in particular for small signal-to-noise ratio.
A further advantage of the maximum-likelihood method is that it also leads to a test specific for model C.

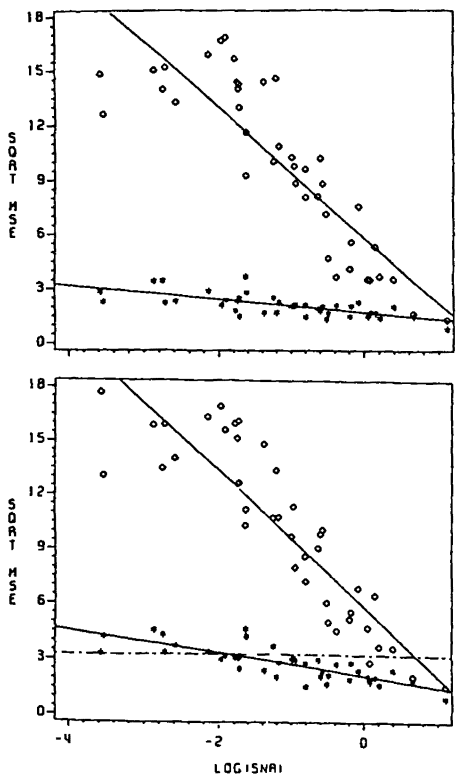

Figure 6.

Estimation of latencies for 41 pseudo real simulations. Root MSE versus log signal-to-noise ratio for ML approach (*) versus Woody's heuristic approach (◇) together with their regression lines. Above : model with no latency jitter; below : model with moderate latency jitter.

5. Concluding Remarks

This paper is intended to alert statisticians of some of the problems specific for the analysis of biomedical signals, but also to point out the scientific interest lying in this field. The problems discussed can be put under the large umbrella of "robustness" - not in its classical sense of deviations from the normal distribution.

Rather, it is shown that a priori models and assumptions should be accepted with some caution, since they can themselves lead to erroneous conclusions. For regression problems an attractive solution via nonparametric estimators is now at hand, and the utility of the associated estimator of derivatives will be better appreciated in the time to come. Regarding deviations from independence, no universal tool is at hand and specific solutions have to be sought, instead of sticking to an assumption that is often not fulfilled.

References

Akaike, M. (1969). Fitting Autoregressive Models for Prediction. *Ann. Inst. Statist., Math.*, 21, 243-247.

Bosq, D. Lecoutre, J.-P. (1987). *Théorie de l'estimation fonctionnelle.* Economica.

Brillinger, D.R. (1975). *Times Series : Data Analysis and Theory.* Holt, Rinehart and Winston, New York.

Brillinger, D.R. (1988). Some Statistical Methods for Random Process Data from Seismology and Neurophysiology. *Ann. Statist.*, 16, 1-54.

Gasser, Th., H.G. Müller, W. Kölher, L. Molinari, A. Prader (1984). Nonparametric Regression Analysis of Growth Curves. *Ann. Statist.* 12, 210-229.

Gasser, Th., H.G. Müller (1984). Estimating Regression Functions and their Derivatives by the Kernel Method. *Scand. J. Statisti.* 11, 171-185.

Gasser, Th. H.G. Müller, W. Köhler, A. Prader, R. Largo, L. Molinari (1985). An anlysis of the mid-growth wpurt and the adolescent growth spurt of height based on acceleration. *Ann. Hum. Biol.*, 12, 129-148.

Gasser, Th., H.G. Müller, V., Mammitzsch (1985). Kernels for Nonparametric curve estimation. *J. Roy. Statist. Soc.* (B), 47, 238-252.

Gasser, Th., L. Sroka, Ch. Jennen-Steinmetz (1986a). Residual Variance and Residual Pattern in nonlinear Regression. *Biometrika*, 73, 625-633.

Gasser, Th., W. Köhler, Ch. Jennen-Steinmetz, L. Sroka (1986 b) The Analysis of Noisy Signals by Nonparametric Smoothing and Differentiation. *IEEE Trans. Biomed. Engin.*, BME-33, 1129-1133.

Gasser, Th., W. Köhler, A. Kneip (1988). A Flexible and Fast Method for Automatic Smoothing and Differentiation. (manuscript).

Härdle, W., P.Hall, J.S. Marron (1988). How far are Automatically Chosen Regression Smoothing Parameters from their Optimum? *J. Americ. Statist. Assoc.*, 83, 86-95.

Hannan, E.J., B.G. Quinn (1979). The Determination of the Order of an Autoregression. *J. Roy . Statist. Soc.* B, 41, 190-195.

Hart, J.D., T.E. Wehrly (1986). Kernel Regression Estimation Using Repeated Measurements Data. *J. Americ. Statist. Assoc.*, 81, 1080-1088.

Hart, J.D. (1987) Kernel Regression Estimation with Time Series Errors. (manuscript).

Jennen-Steinmetz, Ch., Th. Gasser (1988). A Unifying Approach to Nonparametric Regression Estimation. (in print).

Möcks, J., Pham Dinh Tuan, Th. Gasser (1984) Testing for homogeneity of Noisy Signals Evoked by Repeated Stimuli. *Ann. Statist.*, 12, 193-209.

Pham Dinh Tuan, J. Möcks, W. Köhler, Th. Gasser (1987). Variable Latencies of Noisy Signals : Estimation and Testing in Brain Potential Data. *Biometrika*, 74, 525-534.

Silverman, B.W. (1984a). Spline smoothing : the Equivalent Variable Kernel Method. *The Annals of Statistitics*, 12, 898-916.

Steinberg, H.W., Th. Gasser, J. Franke (1985). Fitting Autoregressive Models to EEG Times Series : An Empirical Comparison of Estimates of the Order. *IEEE Trans. Acoustics, Speech and Signal Processing*, ASPP-33, 143-150.

Woody, V.D. (1967). Characterization of an Adaptive Filter for the Analysis of Variable Latency Neuroelectric Signals. *Med. & Biol. Engin.*, 5, 539-553.

Zetterberg, L.H. (1969) Estimation of Parameters for a linear Difference equation with Application to EEG Analysis.,*Math. Biosc.*, 5, 227-275.