Splines in Statistics
Author(s): Edward J. Wegman and Ian W. Wright
Source: *Journal of the American Statistical Association*, Vol. 78, No. 382 (Jun., 1983), pp. 351-365
Published by: American Statistical Association
Stable URL: http://www.jstor.org/stable/2288640
Accessed: 06/03/2011 02:35

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=astata.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

American Statistical Association is collaborating with JSTOR to digitize, preserve and extend access to *Journal of the American Statistical Association.*

http://www.jstor.org

Splines in Statistics

EDWARD J. WEGMAN and IAN W. WRIGHT*

This is a survey article that attempts to synthesize a broad variety of work on splines in statistics. Splines are presented as a nonparametric function estimating technique. After a general introduction to the theory of interpolating and smoothing splines, splines are treated in the nonparametric regression setting. The method of cross-validation for choosing the smoothing parameter is discussed and the general multivariate regression/surface estimation problem is addressed. An extensive discussion of splines as nonparametric density estimators is followed by a discussion of their role in time series analysis. A comparison of the spline and isotonic regression methodologies leads to a formulation of a hybrid estimator. The closing section provides a brief overall summary and formulates a number of open/unsolved problems relating to splines in statistics.

KEY WORDS: Smoothing splines; Functional estimation; Nonparametric regression; Cross-validation; Isotonic estimation.

1. INTRODUCTION

Modern statistical theory began with the fitting of parametric models to data. Various principles for making inferences were developed and refined until efficiencies very nearly reached their asymptotic limits. Such principles include maximum likelihood and likelihood ratio, unbiased and minimum variance estimation, least squares, and more recently decision-theoretic and maximum entropy procedures. As asymptotic limits to efficiency were approached, attention returned to basic models and there has appeared a growing realization that not all the parametric structure was needed to make inferences. From this insight arose the techniques known as distribution-free or nonparametric. At the cost of some loss of efficiency, these methods prevented model violations from being reflected in false inferences.

In a real sense, splines are an evolution of classical parametric inference and bridge the gap between parametric and nonparametric methods. While splines are not parametric in functional form, in most cases they may be written as a linear combination of basis functions that usually have a polynomial representation. Thus there is certainly a parametric flavor. However, the set of admissible functions that may be splines has the cardinality of R^R. Thus there is an extremely rich class of admissible functions with the added benefit of using smoothness properties of increase efficiency.

Our present discussion is organized as follows: Section 2 deals with the fundamentals of interpolating splines. The third section focuses on a general description of smoothing splines. The next four sections focus on splines and regression analysis. The fourth section is a description of regression splines, the fifth section deals with splines as univariate nonparametric regression estimators, and the sixth section deals with the cross-validation method for selection of the smoothing parameter. In Section 7 we discuss nonparametric surface estimation via multivariate splines and in Section 8 we discuss splines as nonparametric density estimators. In Section 9 we discuss the use of splines in time series and in Section 10 we discuss the role of splines in statistical inference under order restrictions. The article closes with a section of general concluding remarks.

We close this section with some notational conventions. We shall reserve the symbol D for the differentiation operator and L_2 for the set of measurable square integrable functions on [0, 1]. The symbol W_m will denote the set of functions, f, on [0, 1] such that $D^j f, j \leq m - 1$, is absolutely continuous and $D^m f$ is in L_2. When we occasionally consider functions with domain other than [0, 1], the relevant domain will be shown after the function space symbol, for example, $W_m(-\infty, \infty)$.

2. CLASSICAL SPLINE THEORY

The dictionary definition of a spline is "a thin strip of wood used in building construction." This in fact gives us insight into the mathematical definition of spline. Historically, engineering draftsmen used long thin strips of wood called *splines*, much like French curves, to draw in a smooth curve between specified points. The splines were anchored in place by attaching lead weights called *ducks* at points along the spline. By altering the position of the ducks and the position of spline and ducks relative

* Edward J. Wegman is Director, Mathematical Sciences Division, Office of Naval Research, 800 N. Quincy Street, Arlington, VA 22217. Ian W. Wright is Lecturer, Mathematics Department, Capricornia Institute of Advanced Education, Rockhampton, Queensland, 4700, Australia. This article formulated with the benefit of conversations and correspondence with many of the leading figures in spline theory. Our thanks go to Professors P. Delfiner, Ron Gallant, Jean Laurent, Dave McAllister, Charles Micchelli, Manny Parzen, and Don Watts. Special thanks go to Patricia Smith and Phil Smith, who were most responsive to our requests for materials. And last but far from least, we would like to express our thanks to Grace Wahba, whose work forms the core of much of the work on splines in statistics, for her good will and encouragement. Earlier drafts of this article were supported by the Air Force Office of Scientific Research, Contract AFOSR-75-2840.

© Journal of the American Statistical Association
June 1983, Volume 78, Number 382
Theory and Methods Section

to the drafting surface, the spline could be made to pass through specified points provided a sufficient number of ducks was used.

If one regards the draftsman's spline as a thin elastic beam, a simple physical demonstration shows that the draftsman's spline is a cantilevered beam that minimizes the energy of deflection subject to the constraint of interpolating the specified points. In the most general setting, then, a mathematical spline is the solution to a constrained optimization problem. In a more elementary setting, the draftsman's spline is replaced by a piecewise cubic polynomial (normally a different one between each pair of ducks) with certain discontinuities permitted where the polynomials join. The piecewise polynomial is chosen to minimize the mean square curvature (corresponding to the deflection energy). The join points in mathematical spline theory are called *knots*.

It should be emphasized that, as presented thus far, the spline is purely interpolatory in nature, and explicit reference to the optimality character does not usually appear in elementary discussions. The interpolation problem is to fit a curve through points (x_i, y_i), $i = 1, 2, \ldots, n$, in the plane. A *mesh* $\Delta = \{\zeta_1 (= x_1) < \zeta_2 < \cdots < \zeta_N (= x_n)\}$ is chosen with the points ζ_i being the knots. For computational reasons, the mesh will frequently coincide with $\{x_i, i = 1, 2, \ldots, n\}$. A *cubic interpolating spline* with mesh Δ, written $s_\Delta(x)$, is a function with continuous derivatives up to and including order 2 that coincides exactly with a possibly different cubic polynomial on each interval $[\zeta_i, \zeta_{i+1}]$, $i = 1, 2, \ldots, N - 1$ and that interpolates $\{(x_i, y_i), i = 1, 2, \ldots, n\}$.

Let the mesh coincide with $\{x_i, i = 1, 2, \ldots, n\}$ and let $h_i = x_{i+1} - x_i$ and $M_i = s_\Delta''(x_i)$, $i = 1, 2, \ldots, n$. Suppose the polynomial interpolating (x_i, y_i) and (x_{i+1}, y_{i+1}) is as follows:

$$y = a_i (x - x_i)^3 + b_i (x - x_i)^2 + c (x - x_i) + d_i. \quad (2.1)$$

By taking various order derivatives and evaluating at the knot points, it may be shown that

$$b_i = M_i/2$$
$$a_i = (M_{i+1} - M_i)/6h_i$$
$$c_i = \frac{y_{i+1} - y_i}{h_i} - \frac{2(h_i M_i + h_i M_{i+1})}{6}$$
$$d_i = y_i. \quad (2.2)$$

Thus our curve fitting problem reduces to that of finding the values of M_i. The equations relating the M_i are obtained by using the continuity of the first derivative of the spline, along with the relations 2.2 to give

$$h_{i-1} M_{i-1} + (2h_{i-1} + 2h_i) M_i + h_i M_{i+1}$$
$$= 6 \left(\frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}} \right), \; i = 2, 3, \ldots, n - 1.$$

A requirement that $M_1 = M_n = 0$ leads to a tridiagonal system of linear equations for $M_2, \ldots, M_{n-1}$. This system can easily be solved by Gaussian elimination so that fitting a cubic interpolating spline is feasible with a hand calculator.

As mentioned earlier, it may be shown that the cubic interpolating spline is the solution to the following problem:

$$\begin{aligned} &\text{minimize} && \int_{-\infty}^{\infty} (D^2 f(x))^2 dx \\ &\text{subject to} && D^j f \in L_2(-\infty, \infty), \quad j = 0, 1, 2 \\ &\text{and} && f(x_i) = y_i, \qquad i = 1, 2, \ldots, n. \end{aligned} \quad (2.3)$$

This simple problem can be generalized. Let L be a differential operator of order m with constant coefficients. The following problem,

$$\begin{aligned} &\text{minimize} && \int_{-\infty}^{\infty} (L f(x))^2 dx \\ &\text{subject to} && D^j f \in L_2(-\infty, \infty), \quad j = 0, 1, \ldots, m \\ &\text{and} && f(x_i) = y_i, \qquad i = 1, 2, \ldots, n, \end{aligned} \quad (2.4)$$

has a solution $s(x)$ that satisfies $L^*L\, s(x) = 0$ in the in tervals between knot points, where L^* is the adjoint operator to L. The solution, $s(x)$, is called the *interpolating L spline*. If, in fact, $L = D^m$, then $s(x)$ must satisfy D^{2m} $s(x) = 0$ so that $s(x)$ is a piecewise polynomial of order $2m - 1$. In the special case $L = D^2$, then $s(x)$ is a piecewise polynomial of order $2m - 1 = 3$. It is to be emphasized, therefore, that the polynomial character of splines is a result of the choice of the operator L and that the concept of a spline is more general than just a smoothly joined piecewise polynomial.

General accounts of interpolating splines may be found in the books by Ahlberg, Nilson, and Walsh (1967), Greville (1969), and Prenter (1975). The literature in this area is vast, but representative papers include Anselone and Laurent (1968), Copley and Schumaker (1978), Daniel and Schumaker (1974), and Mangasarian and Schumaker (1969). The latter paper is found in a conference proceedings, Schoenberg (1969), which is one of the most useful such volumes published. The paper by Schultz and Varga (1967) contains an extensive bibliography on L splines. An excellent resource for implementing splines is De Boor (1978).

3. SMOOTHING SPLINE THEORY

Interpolating splines are predicated on nonnoisy data. As such they have limited use in a statistical setting, although in several circumstances they do make an appearance. More to the point, it is desirable in a statistical framework to create a type of smoothing spline that could pass near, in some sense, to the data but not be constrained to interpolate exactly. There are three main approaches to spline fitting methods corresponding to different points of view in dealing with the "noise" in the data. Because stochastic data do not constrain the fitted function nearly as firmly as in the interpolating spline case, the fitting sometimes requires a genuine optimization routine, not just a simple solution of a linear system of equations as with a cubic interpolating spline.

The first, more frequently used method parallels the least squares curve-fitting procedure by minimizing a criterion that depends on a least-squares-like term plus a term penalizing roughness. This method is an appropriate one to use when the error shocks have an infinite or semi-infinite support. In contrast, when data are, for example, direct readings from a calibrated instrument, it is sometimes possible to set fairly narrow 100 percent confidence limits for each data point. The second method is used in this circumstance. The third method is also a least squares procedure, but takes a somewhat different perspective. We mentioned earlier that the polynomial character of the spline was a result of the formulation of the optimization problem. In the third method one assumes a piecewise polynomial form for the splines and thence uses a least squares procedure to estimate the appropriate parameters. This approach, while somewhat less organic, has received a fair amount of attention in the statistics literature. We discuss all three approaches in some detail.

3.1 First Method of Fitting Smoothing Splines: Penalized Least Squares

Suppose that the x values of the data lie in a finite interval. Without any loss of generality, we assume the interval is $[0, 1]$ and that we have $0 < x_1 < x_2 < \cdots < x_n < 1$. The fitted spline is the solution to the optimization problem:

$$\text{Minimize} \quad \sum_{i=1}^{n} (f(x_i) - y_i)^2 + \lambda \int_0^1 (L\, f(x))^2 dx$$

$$\text{subject to} \quad f \in W_m,\ \lambda \quad \text{fixed} \quad > 0. \tag{3.1}$$

This is clearly a generalization of (2.4) with the interpolating conditions, $f(x_i) = y_i$, replaced with a least squares term in the objective function. The least squares term is augmented by the "curvature" term, $\int_0^1 (L f(x))^2 dx$. This latter term is, as before, a penalty term for lack of smoothness. Notice that the parameter $\lambda > 0$ in 3.1 controls the amount of smoothing. If λ is too small, the spline will overfit, in the limiting case as $\lambda \to 0$, becoming an interpolating spline. This lowers bias, but increases variance. As $\lambda \to \infty$, the smoothing term dominates and removes not only noise but also "signal." The correct choice of λ is of considerable importance. As sample size n approaches ∞, λ should become smaller. Asymptotic rates for λ to approach 0 needed to guarantee consistency have been given by several authors. This will be discussed later. The method of cross-validation for choosing λ has also been offered as an option for choosing λ. This, also, will be discussed in a later section.

In many settings, the choice of $L \equiv D^m$ is appropriate. In such a case, the solution is given explicitly in Kimeldorf and Wahba (1970a,b) and as before it turns out to be a polynomial spline of degree $2m - 1$ with possible knots at the data points. This characterizing solution does not turn out to be a particularly useful computational algorithm. Cogburn and Davis (1974) demonstrate somewhat easier computational algorithms for the case in which the x values are evenly spaced and f is periodic. In other circumstances, more difficult algorithms must be developed. Questions of the computation of such splines can be approached by quadratic programming algorithms. The work of Kimeldorf and Wahba (1971) is of prime interest here. Also of interest are the papers of Ritter (1969), Anselone and Laurent (1968), and Wahba (1978a).

3.2 Second Method of Fitting Smooth Splines: 100 Percent Confidence Intervals

The penalized least squares splines were a generalized version of the interpolating splines because the interpolating constraints were replaced by a least squares term in the objective function. An alternative approach is to leave the objective function untouched but loosen the specification of the interpolating constraints. This fitting technique is also known as (the solution for) the Generalized Hermite-Birkhoff (GHB) interpolation problem.

Let $[\alpha_i, \beta_i]$ be a 100 percent confidence interval for the ordinate at x_i with $\alpha_i < \beta_i$. The GHB problem is as follows:

$$\text{Minimize} \quad \int_0^1 (Lf(x))^2 dx$$

$$\text{subject to} \quad f \in W_m,\ \alpha_i \le f(t_i) \le \beta_i,$$

$$i = 1, 2, \ldots, n, \tag{3.2}$$

The statistical interpretation can be seen by considering the following model

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{3.3}$$

If ϵ_i are iid with finite support, say $[-e_1, e_2]$, then since $\epsilon_i > -e_1$, it is clear that $y_i + e_1 > y_i - \epsilon_i = f(x_i)$. Similarly, since $\epsilon_i < e_2$, $y_i - e_2 < y_i - \epsilon_i = f(x_i)$. Thus $(y_i - e_2, y_i + e_1)$ is a 100 percent confidence interval and we may identify α_i with $y_i - e_2$ and β_i with $y_i + e_1$.

Various recent contributions have been made by several authors. Atteia (1968) demonstrates the existence of such splines while Laurent (1969) gives a characterization theorem. Once again the solution to (3.2) is a polynomial spline of degree $2m - 1$ with knots at those data points where the constraints are active. Ritter (1969) and others discuss computational algorithms.

3.3 Third Method of Fitting Smoothing Splines: Regression Splines

While the piecewise polynomial result falls out of the optimization problems (3.1) and (3.2) yet another approach to splines is to assume the form of a smoothly joined piecewise polynomial of degree m. The pieces join smoothly and fulfill continuity conditions on the function and the first $m - 1$ derivatives. This type of spline is thus a continuous function with $m - 1$ continuous derivatives.

Obviously while it satisfies the continuity conditions similar to our previous splines, it does not necessarily minimize the curvature norm. Since there are obviously many such polynomial splines which could satisfy the other requirements, there are in general a number of free parameters which must be determined. The parameters at the user's disposal are

1. The degree of the spline function, m.
2. The number of knots, N.
3. The positions of the knots, ζ_i.
4. The free coefficients in the spline function, $m + N + 1$ in number.

The degree and number of knots are usually fixed by the experimenter. The number of knots in (3.1) and (3.2) is usually the same as the number of observations, while in this third case, however, the number of knots is considerably smaller. The knots may be either fixed or free. In the latter case the locations, ζ_i, must also be estimated. The free coefficients of the spline function are those polynomial coefficients left over after the continuity conditions have been satisfied. These free parameters then may be estimated by an ordinary least squares procedure to uniquely identify the spline function. Review papers on this third type of spline are found in Wold (1974) and Smith (1979). We discuss these splines in more detail in the next section.

4. REGRESSION SPLINES

Regression splines (as opposed to spline regressions) are splines that have been computed according to a regression model. In particular, all splines discussed in this section will follow the model

$$y_i = s_\Delta(x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \tag{4.1}$$

where $s_\Delta(x)$ is a piecewise polynomial spline with mesh Δ and the ϵ_i are a white noise sequence. The problem then is to compute $s_\Delta(x)$ based on a least squares approach. Given the mesh of knots, $\Delta = \{\zeta_1 < \zeta_2 < \cdots < \zeta_N\}$, Smith (1979) shows (see also Greville 1969) that a piecewise or segmented polynomial may be represented in the Heaviside function notation; that is, if we let $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leq 0$, then the general form of a segmented polynomial regression model, (4.1), with N knots in the mesh Δ having $N + 1$ polynomial pieces each of degree m may be rewritten as follows:

$$y_i = \sum_{j=0}^{m} \beta_{0j}\, x_i^j + \sum_{k=1}^{N} \sum_{j=0}^{m} \beta_{kj}\, (x_i - \zeta_k)_+^{\,j} + \epsilon_i. \tag{4.2}$$

As we shall see later, the polynomial spline is a special case of this model. This representation is clearly a very useful one since it casts the spline (or segmented polynomial) problem into an ordinary multiple regression context. In fact, the coefficients, β_{kj}, may be determined by ordinary least squares routines commonly available on statistical computing packages. Moreover, many of the $F-$, $t-$, and nonparametric test procedures for determining whether the β_{kj} are 0 carry over directly.

The issue of whether $\beta_{kj} = 0$ is of some substantial interest. The presence of the term $\beta_{kj}\,(x - \zeta_k)_+^{\,j}$ allows a discontinuity at ζ_k in the jth derivative of s_Δ and its absence forces continuity of $D^j s_\Delta$ at ζ_k. Thus, s_Δ can be made continuous at ζ_k by omitting from (4.2) the term $\beta_{k0}(x - \zeta_k)_+^{\,0}$ and similarly $D^j s_\Delta$ can be made continuous at ζ_k by omitting $\beta_{kj}(x - \zeta_k)_+^{\,j}$. Different sorts of continuity conditions can be made to hold at different knots by imposing conditions that selected β_{kj} be 0. The classic spline of degree m, which requires continuity up to and including $D^{m-1}s_\Delta$ has the representation,

$$y_i = \sum_{j=0}^{m} \beta_{0j}\, x_i^j + \sum_{k=1}^{N} \beta_{km}\, (x_i - \zeta_k)_+^{\,m} + \epsilon_i; \tag{4.3}$$

thus the number of free coefficients in the spline to be estimated is $m + N + 1$ as mentioned earlier. The treatment of Smith (1979) has an excellent discussion of model (4.2), in general, and of testing for $\beta_{kj} = 0$ in particular. We strongly recommend the Smith paper for someone interested in applying regression splines.

The paper by Wold (1974) reflects a lot of experience with fitting regression splines. Based on his practical experience Wold made some useful recommendations for knot point selection, which we summarize below. These recommendations are based on the assumption of fitting a cubic spline, the most popular case, and may need some modification for $m > 3$.

1. Knot points should be located at data points.
2. A minimum of four or five observations should fall between knot points.
3. No more than one extremum and one inflection point should fall between knots (because a cubic could not fit more).
4. Extrema should be centered in intervals and inflection points should be located near knot points.

The papers of Smith (1979) and Wold (1974) address the major choices, listed in our Section 3.3; that is, the estimation of the free coefficients and the number and position of knots. The degree of the spline function, m, depends on what is a realistic assessment of the number of derivatives available in the regression function. Obviously, this knowledge is frequently not available. Often the choice is simply $m = 3$, which yields a cubic spline and which is the smallest m yielding visual smoothness. The choice $m = 2$ or $m = 1$ will yield, respectively, piecewise quadratics or piecewise lines (i.e., quadratic and linear splines). There are a number of papers that address these cases as well. See, for example, Agarwal and Studden (1978), Ertel and Fowlkes (1976), Derek (1966), Park (1978), Gallant and Fuller (1973), and Fuller (1969).

A problem of regression closely related to spline regression is the design problem. At each x_i, $i = 1, 2, \ldots, n$, n_i observations are to be taken. The probability measure assigning mass λ_i to point x_i is referred to as a design measure. Even more generally, the design problem is to choose the locations of the x_i. Several authors have studied the problem of finding optimal designs for regression splines including Agarwal and Studden (1978), Studden and Van Arman (1969), Park (1978), Draper, Guttman, and Lipow (1977), and Murty (1971).

Other papers of considerable interest are those of Poirer (1973), Gallant and Fuller (1973), Buse and Lim (1977), and Lenth (1977). Poirer has an excellent discussion of the basic theory of cubic regression splines. Arguing from an economic point of view, he develops the idea that structural change occurs in a smooth fashion so that splines form a natural tool for analyzing structural changes. He argues that knots should occur at a point in time near the point of structural change. Poirer illustrates his point with an example based on Indianapolis 500 data. Buse and Lim (1977) follow up Poirer's paper, arguing that cubic spline regression is a special case of restricted least squares and that the latter approach offers a richer menu of procedures. Gallant and Fuller (1973) treat the join points (knots) as unknown parameters and develop procedures for estimating their location. Finally, Anderssen, Bloomfield, and MacNeil (1974) argue that the spline procedures can be robustified (against heavy-tailed errors, ϵ_i) by using the M estimate procedures of Huber (1964) instead of the usual sum of squares. Lenth (1977), apparently independently of Anderssen, Bloomfield, and MacNeil, puts forward a similar argument.

5. NONPARAMETRIC REGRESSION

Whereas in the previous section we were concerned with the use of regression models and least squares to fit splines, it is equally clear that in the most general setting, the spline solutions to (3.1) and (3.2) are solutions to the nonparametric regression problem or more prosaically, the curve fitting problem.

Statisticians and applied mathematicians are continually faced with the problem of recovering a smooth function when only noise measurements of it are available. In fitting a parametric model, the residuals are made up of the noise as well as deviations due to lack of fit. Penalized least squares smoothing splines are good solutions to the estimation of the true function (known only to be smooth) for two reasons. First, they are flexible enough to respond to local variation without allowing pathological behavior, and second, the actual degree of smoothing is controllable. Even when the correct degree of smoothing is unknown, these features together with the method of cross-validation allow a nonparametric, yet optimal, fit.

The model we assume is

$$y_i = f(x_i) + \epsilon_i, \quad x_i \in [0, 1], \quad i = 1, 2, \ldots, n.$$

where

$$f \in W_m, \; E\epsilon_i = 0$$

and

$$\begin{aligned} E\epsilon_i\epsilon_j &= \sigma^2 \quad i = j \\ &= 0 \quad i \neq j. \end{aligned} \tag{5.1}$$

The solution, s_λ, to the optimization problem (3.1) with $L \equiv D^m$ serves as an estimator for f in (5.1). If $m = 2$, the optimal solution (Greville 1969; Reinsch 1967,1971) is known to be a cubic spline with knots at x_i, $i = 1, 2, \ldots, n$. As $\lambda \to \infty$, the solution, $s_\lambda(x)$, converges to its smoothest possible form, the least squares straight line through the data. As $\lambda \to 0$, $s_\lambda(x)$ converges to the interpolating spline through all of the data points. Thus λ is a parameter determining the degree of smoothing. Wahba (1975c) shows that in order to have $s_\lambda \to f$ as $n \to \infty$ we must also have $\lambda \to 0$. These papers by Wahba are sources of other asymptotic results as well.

We have already pointed out in Section 3 that the 100 percent confidence interval method for fitting smoothing splines corresponds to the nonparametric regression problem (3.3), in which the errors have bounded support. This is in contrast to the penalized least squares case, which corresponds regression with "normal-like" (unbounded) errors. Generally speaking, results on these two types of smoothing splines can be thought of as falling into three major classes: (a) Results on existence and characterizations, (b) results on statistical aspects including asymptotics, and (c) results on computational aspects. Important papers in the first category include papers of Atteia (1968,1970), which give existence results, and the work of Laurent (1972) and Anselone and Laurent (1968), which gives results characterizing solutions to (3.1) and (3.2) with $L \equiv D^m$ as $2m - 1$ degree polynomial splines.

There are several closely related papers including those of Mangasarian and Schumaker (1969), Daniel and Schumaker (1974), and Copley and Schumaker (1978), which consider the related optimization problem

$$\text{Minimize} \quad \int_0^1 (Lf(x))^2 dx \quad \text{subject to}$$

$$\text{(i)} \qquad f \in W_m$$

$$\text{(ii)} \qquad \alpha_i \leq Ff(x_i) \leq \beta_i, \quad i = 1, 2, \ldots, r, \tag{5.2}$$

where F is a bounded linear functional on W_m. Copley and Schumaker give existence and characterization results. The related paper of Wahba (1973) considers finite constraints and investigates conditions for convergence. Generally speaking, results on existence and characterizations are function theoretic in character frequently appealing to the theory of reproducing kernel Hilbert spaces. See Aronszajn (1950) and Parzen (1961). We shall not attempt to detail these mathematical foundations in this article, but we do remark that the group at the University of Grenoble including Professors Laurent, Atteia, Duchon and their students, and the group at the University of Wisconsin-Madison, notably Professor Wahba, have been very productive in this area. See also the work of Speckman (1983).

Statistical interpretation of splines as nonparametric regression estimators have received somewhat less attention. Gamber (1979b), for example, discusses penalized least squares splines as the basis for generating confidence regions. Wegman (1980b) discusses the relationship of both penalized least squares splines and 100 percent confidence interval splines to nonparametric and isotonic regression and gives consistency results for 100 percent confidence interval splines. The related work of Clark (1977) discusses some solutions to the nonparametric regression problem including the penalized least squares smoothing spline.

Clark and also Schoenberg (1964) point out that the spline solution s_λ to (3.1) has the property that it minimizes

$$\int_0^1 (D^m f(x))^2 dx$$

subject to

$$\sum_{i=1}^{n} (y_i - f(x_i))^2 \leq U_0,$$

and similarly it minimizes

$$\sum_{i=1}^{n} (y_i - f(x_i))^2$$

subject to

$$\int_0^1 (D^m f(x))^2 dx \leq g_0 \tag{5.3}$$

for some λ depending on U_0 and g_0, respectively. This latter condition establishes connection with the least squares regression splines discussed in Section 4. This follows since the least squares polynomial spline will satisfy (5.3) for some g_0 and hence (provided m and the knots are chosen properly) will be a penalized least squares spline for some λ depending on g_0.

Questions of the computation of the penalized least squares and confidence interval smoothing splines may be addressed through a quadratic programming approach. The work of prime interest here is the paper of Kimeldorf and Wahba (1971), but also of interest are the papers of Ritter (1969), Anselone and Laurent (1968), Amos and Slater (1969), Lyche and Schumaker (1973) and Wahba (1977a, 1978a). Kimeldorf and Wahba (1971) give explicit although rather complicated algorithms for constructing both interpolating and smoothing splines. In fact they give not only algorithms, but a general approach based on reproducing kernel Hilbert spaces, for developing such an algorithm. In Section 6 of their paper, they show that problems of the type (3.2) with linear inequality constraints may be solved as a quadratic programming problem. Their set of basis function is not particularly easy to use computationally and a more favorable set of basis functions is found in Wahba (1978a). As mentioned in Section 3, Cogburn and Davis (1974) demonstrate easier computational algorithms in the periodic spline case. Cogburn and Davis will be treated in more detail in Section 9.

The books by Schoenberg (1969) and Karlin et al. (1976) are particularly relevant to the spline types discussed in this section. Finally we note an excellent general discussion of optimal curve fitting by Weinert (1980), which unfortunately is published in an out-of-the-way place. Weinert's paper expands the treatment given in the present section and is commended to the attention of readers particularly interested in problems of nonparametric regression or curve fitting.

6. CROSS-VALIDATION

In the penalized least squares method for fitting smoothing splines (i.e., problem (3.1)), the choice of smoothing parameter, λ, is of substantial importance. To a lesser extent the choice of m as a free parameter determines the final appearance of the smoothing spline. The method of cross-validation has been advocated for choosing λ (and m) by Wahba and Wold (1975a,b), Wahba (1976,1979c,1980b), Golub, Heath, and Wahba (1979), Craven and Wahba (1979), Wahba and Wendelberger (1980), Gamber (1979a) and Utreras-Dias (1979). In an excellent expository paper, Wahba (1979b) discusses the method of (generalized) cross-validation for choosing λ. We summarize that discussion here and recommend the full paper for details.

The parameter, λ, to be chosen is the smoothing parameter in problem (3.1) with L chosen as D^m. We let $s_\lambda^{(k)}$ be the solution to problem (3.1) with the kth data point, (x_k, y_k), omitted. Of course, $s_\lambda^{(k)}(x_k)$ is an estimator of y_k and λ is appropriately chosen if $s_\lambda^{(k)}(x_k)$ is a good estimator of y_k. To measure goodness of fit, we choose the average squared error; that is,

$$\mathrm{CV}(\lambda) = \frac{1}{n} \sum_{k=1}^{n} (s_\lambda^{(k)}(x_k) - y_k)^2,$$

which is called the cross-validation function. The parameter λ is chosen to minimize $\mathrm{CV}(\lambda)$.

For certain technical reasons, it is desirable to compute a weighted least squares cross-validation function,

$$\mathrm{GCV}(\lambda) = \frac{1}{n} \sum_{k=1}^{n} (s_\lambda^{(k)}(x_k) - y_k)^2 w_k(\lambda)$$

where the $\{w_k(\lambda)\}$ are weights chosen to reflect unequally spaced data, end effects and other effects. The papers by Golub, Heath, and Wahba (1979) and Craven and Wahba (1979) detail the choice of $w_k(\lambda)$. This function, called the generalized cross-validation function, has a particularly simple matrix representation:

$$\mathrm{GCV}(\lambda) = \frac{\frac{1}{n} \left\| (I - A(\lambda))\mathbf{y} \right\|^2}{\left(\frac{1}{n} \mathrm{Trace}\,(I - A(\lambda))\right)^2}$$

where $A(\lambda)$ is the $n \times n$ matrix uniquely determined by

$$\begin{bmatrix} s_\lambda(x_i) \\ \vdots \\ s_\lambda(x_i) \end{bmatrix} = A(\lambda)\mathbf{y}.$$

The GCV estimate of λ is obviously the choice of λ minimizing GCV(λ).

The generalized cross-validation function resembles in some sense the mean squared error, and one may suspect some asymptotic relation. Indeed Craven and Wahba (see also Golub, Heath, and Wahba), give the following asymptotic result. If

$$\text{MSE}(\lambda) = \frac{1}{n} \sum_{k=1}^{n} (s_\lambda(x_k) - f(x_k))^2$$

where f is the true function being estimated in (5.1), then both MSE(λ) and GCV(λ) can be regarded as random functions of the $\{\epsilon_i\}$. If λ^* is the minimizer of expected MSE(λ), EMSE(λ) and λ^+ is the minimizer of expected GCV(λ), EGCV(λ), then

$$\lim_{n\to\infty} \frac{\text{EMSE}(\lambda^+)}{\text{EMSE}(\lambda^*)} \downarrow 1,$$

or the mean squared error with estimated λ tends in expectation to the minimum mean squared error achievable with any λ.

The choice of m by means of cross-validation has more recently received attention. See Gamber (1979a). Transportable computer code does not yet exist for this latter problem, but at least three sources (Fleisher 1979, Merz 1978, and Paihua-Montes 1979) provide a computer code for the computation of penalized least squares smoothing splines using generalized cross-validation.

7. SURFACE ESTIMATION AND KRIGING

The nonparametric regression problem formulated as (5.1) can be extended to the more general setting

$$z_i = L_i f + \epsilon_i, \quad i = 1, 2, \ldots, n \qquad (7.1)$$

where the $[\epsilon_i]$ are independent, zero mean random variables with common unknown variance σ^2. The L_i are assumed to be continuous linear functionals on W_m. In the regression setting (5.1), $L_i f$ is simply defined as $f(x_i)$. Other cases of interest are

$$L_i f = \int K(t_i, s)\, f(s) ds,$$

which is applicable for the solution of ill-posed problems (see Wahba 1980) and

$$L_i f = \int_{A_i} u(s) ds$$

where the data functionals are regional averages (see Wahba and Dyn 1982). Perhaps the most interesting application from the statistical point of view is the estimation of (hyper) surfaces. Let f be a smooth function on a closed and bounded subset, say Ω, of d-dimensional Euclidean space, R^d. We choose $L_i f = f(t_i)$ where $t_i \in \Omega$, $i = 1, 2, \ldots, n$. Here t is a d-dimensional vector. The $(d + 1)$-dimensional vector, (z_i, t_i) is observed and we desire to estimate f, nonparametrically from (z_i, t_i), $i = 1, 2, \ldots, n$. In analogy with (3.1), our smoothing spline is the solution to the problem:

$$\text{Minimize} \quad \frac{1}{n} \sum_{j=1}^{n} (f(t_j) - z_j)^2 + \lambda J_m(f)$$
$$\text{subject to} \quad f \in W_m^*, \quad \lambda > 0. \qquad (7.2)$$

Here W_m^* is the obvious generalization of W_m to d-dimensions and $J_m(f)$ is chosen as

$$J_m(f) = \sum_{i_1,\ldots,i_d=1}^{m} \cdots \sum^{m} \int_{R^d} \left(\frac{\partial^m f}{\partial x_{i_1} \cdots \partial x_{i_d}} \right)^2 dx_1 \cdots dx_d. \qquad (7.3)$$

In the particular case, $d = m = 2$, the smoothness penalty team becomes

$$J_2(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\partial^2 f}{\partial x_1^2} \frac{2\partial^2 f}{\partial x_1 \partial x_2} + \frac{\partial^2 f}{\partial x_2^2} \right) dx_1 dx_2,$$

which corresponds to the bending energy of a thin plate (the two-dimensional analog of a thin elastic beam). For this reason, the general class of solutions to (7.2) is known as the set of thin-plate smoothing splines. Duchon (1976a,b) has given an explicit representation for the solution to (7.2) and related work of Meinguet (1978,1979) has further characterized these solutions in a reproducing kernel Hilbert space setting. A more easily accessible description and proof of these characterization results may be found in the appendix of Wahba (1979b).

Wahba (1979b) is also an excellent source for the description the generalized cross-validation method for estimating λ in the multidimensional setting of (7.3). In particular the algorithms for $d = 2$ are described and some numerical results are described. In a related paper, Wahba (1980b) discusses two-dimensional thin-plate splines and gives some interesting meteorological examples. Wahba (1979a,1981) discusses convergence rates for multidimensional splines. In particular, the earlier study contains some heuristic arguments and conjectures, which suggests that thin-plate splines using generalized cross-validation can achieve mean squared error convergence rates comparable to the best obtainable rates in a nonparametric regression setting. See Stone (1980,1982). Following the meteorological motivation further, Wahba (1981) discusses splines on the surface of a sphere. Solutions to the analog of (7.2) on a sphere have an infinite series representation that is computationally inconvenient. Wahba also proposes an alternate quadratic functional for $J_m(f)$ which gives rise to a practical pseudo-spline on the sphere.

Geological, specifically mine engineering, applications have inspired techniques closely related to thin-plate splines. In attempting to estimate the quantity of ore in a deposit, core samples are taken at various points and surfaces are estimated. Estimates of the top and bottom ore surfaces are made using a certain form of minimum variance unbiased estimation on homogeneous random fields, see Matheron (1973). Two-dimensional interpolation and smoothing methods based on Matheron's homogeneous random fields are called *kriging* in the mining industry. Delfiner (1978) discusses the kriging estimates. By comparing the estimates in Delfiner (1978) and Duchon (1976a), one can see that the kriging estimates, loosely speaking, solve the minimization problem (7.2) with $d = 2$ and $m = 3/2, 5/2, \ldots$. The main difference between kriging as described by Delfiner (1978) and the thin-plate splines is the choice of λ and m and how they are estimated. The books, Guarascio, David, and Huijbregts (1976) and Rendu (1978) contain more extended discussions of kriging. In particular, D.G. Krige offers a historical perspective in Guarascio, David, and Huijbregts.

A somewhat different approach to surface estimation using splines is found in Friedman, Grosse, and Stuetzle (1980). The approach is based on projection pursuit algorithms (see also Friedman and Stuetzle 1981). An optimal one-dimensional projection of the explanatory variable is sought and, using this one-dimensional projection, an ordinary one-dimensional spline fit is made. The projection is optimal in some sense such as maximum explained variance by the one-dimensional spline fit. The residuals from this optimal are then treated as a new data set and a second projection is sought. The second spline fit is added to the first and the residuals from this sum are treated as a third data set. The process is repeated until some appropriate stopping criterion is met. The resulting sum of splines is then the fit to the surface. This approach, while not optimal in the sense of (7.2), has the advantage of avoiding the "curse of dimensionality" in the sense that only one-dimensional splines are being fitted in every case.

We close this section by noting that Wegman (1981) discusses vector-valued splines, that is, the case in which the range (as opposed to the domain) of f is R^d. So far as we know, no one has addressed the problem of estimating f when both range and domain are higher-dimensional (or non-Euclidean).

8. SPLINE DENSITY ESTIMATION

The nonparametric estimation of probability densities has received considerable attention in the last 25 years. See, for example, Rosenblatt (1971) or Wegman (1972a,b) for expository work in this area. Perhaps the most innovative development in nonparametric density estimation since those papers has been the development of spline methods. Three essentially independent efforts occurred in 1971: the works of Boneva, Kendall, and Stefanov (1971), Wahba (1971), and Good and Gaskins (1971). We shall discuss each of these papers and their spinoffs in turn. In our discussions we will confine ourselves to density estimation based on an independent identically distributed sample. Spline estimation provides a very satisfying method for density estimation because of the following result:

The solution f of the problem

$$\text{Minimize} \quad \int_0^1 (D^m f(x))^2 dx \quad \text{with } f \in W_m$$

$$\text{and} \quad f(x_i) = y_i, \quad i = 1, 2, \ldots, n$$

and the solution g of the problem

$$\text{Minimize} \quad \int_0^1 (D^{m-1} g(x))^2 dx \quad \text{with } g \in W_{m-1}$$

$$\text{and} \quad (D^{-1}g)(x_i) = y_i, \quad i = 1, 2, \ldots, n$$

are related by $Df = g$.

This means an interpolating spline-fitted density may be obtained by differentiating the interpolating spline-fitted distribution.

In an invited paper with discussion, Boneva, Kendall, and Stefanov (1971) laid out the fundamental theory of their histosplines, empirical densities that are smooth analogs of a histogram based on interpolating splines.

Although the theory of histosplines is somewhat involved, a histospline is, in essence, an interpolating spline fitted to the usual histogram. Rather than viewing the histogram as a function on the real line, however, it is regarded as h, a sequence on the integers. More details may be found in Boneva, Kendall and Stefanov, whose paper also describes much empirical material on histospline behavior.

It must be emphasized that since histosplines are interpolating splines based on the sample histogram, and not smoothing splines, we cannot expect this method in the presence of noise (sampling error) to be much better at filtering the noise than the histogram from which it is derived.

This assertion is supported by the results of Wahba (1975b) who shows for her variant of the histospline that for the true density $f \in W_m$ and f_n the histospline corresponding to a sample of size n

$$E(f_n(x) - f(x))^2 = O\,(n^{-(2m-1)/2m}). \tag{8.1}$$

In a companion paper Wahba (1975a) shows that the expected mean square error at a point t has that same order of magnitude for all of the following estimation methods: the polynomial algorithm (Wahba 1971), kernel type estimation, certain orthogonal series estimates and the ordinary histogram. However, the constants covered by the O may be larger in these latter cases.

The polynomial algorithm mentioned above was described in Wahba (1971), the second of the 1971 papers. An algorithm is described based on ordinary polynomial (Lagrange) interpolation. The empirical distribution is locally interpolated by an mth-degree polynomial passing through the empirical distribution evaluated at $m + 1$ adjacent order statistics. The density is the derivative of this interpolating polynomial. Mean squared error convergence rates are then discussed as above in (8.1).

The third of this trio of landmark 1971 papers was that of Good and Gaskins (1971), a paper on maximum penalized likelihood estimators (MPLE). To consider this in detail we let $H(a, b)$ be a manifold in $L_1(a, b)$. (Manifold = set of reasonably similar functions.) Suppose $x_1, \ldots, x_n$ is an iid sample from an unknown density $f \in L_1(a, b)$. Unfortunately, the following problem

$$\text{Maximize} \quad L(f) = \prod_{i=1}^{n} f(x_i) \quad \text{subject to}$$

$$\text{(i)} \quad f \in H(a, b)$$

$$\text{(ii)} \quad \int_a^b f(x)dx = 1$$

$$\text{(iii)} \quad f(x) \geq 0 \quad \text{for all } x \in (a, b),$$

will not have a solution for most manifolds of interest (the unimodal or monotone functions are an exception). Specifically, any manifold that contains an approximating sequence to any linear combination of δ functions has no maximum likelihood estimator for the density f.

From heuristic Bayesian considerations, Good and Gaskins (1971) suggested adding a penalty term to the likelihood which would penalize unsmooth estimates. They chose a manifold and penalty function that leads to polynomial splines. Good's and Gaskin's results were refined and made rigorous by de Montricher, Tapia, and Thompson (1975). We can now describe the current state of the art.

It will normally be the case that the manifold $H(a, b)$ is contained in $W_m(a, b)$ and that the penalty function $\Phi(f) = \int_a^b (D^m f(x))^2 dx$. Let

$$\hat{L}(f) = \prod_{i=1}^{n} f(x_i)\exp(-\Phi(f))$$

and consider the following optimization problem:

$$\text{Maximize} \quad \hat{L}(f) \quad \text{subject to}$$

$$\text{(i)} \quad f \in H(a, b)$$

$$\text{(ii)} \quad \int_a^b f(x)dx = 1$$

$$\text{(iii)} \quad f(x) \geq 0, \quad x \in (a, b) \tag{8.2}$$

The solution to (8.1) is the MPLE of the underlying density.

The task of computing the MPL Estimate of the density is greatly simplified by knowing the form the optimum must take. The following existence theorem is proved in the paper by de Montricher, Tapia, and Thompson (1975).

Theorem. For $m \geq 1$, the MPLE corresponding to W_m exists, is unique, and is a polynomial spline of degree $2m - 1$. Moreover, if the estimate is positive in the interior of an interval, then in this interval it is of degree $2m - 1$ and of continuity class $2m - 2$ with knots at the sample points.

From (Fisher) information-theoretic considerations, as well as a desire to avoid the awkward nonnegativity constraint $f(x) \geq 0$, Good and Gaskins (1971) also considered the MPLE problem with manifold

$$H_1(-\infty, \infty) = \{f\colon f^{1/2} \in W_1(-\infty, \infty)\}$$

$$\Phi_1(f) = \alpha \int_{-\infty}^{\infty} \frac{(Df(x))^2}{f(x)}\, dx$$

$$= 4\,\alpha \int_{-\infty}^{\infty} (Df(x)^{1/2})^2 dx, \ \alpha > 0, \tag{8.3}$$

where $f = (f^{1/2})^2$ is to be the (necessarily positive) density. After noting that the reformulation trick (8.3) is standard in the literature, de Montricher, Tapia, and Thompson (1975) record conditions for its valid use. The authors go on to establish that the price of using the nonnegativity trick is to lose the polynomial spline form of solution, the solution being an exponential spline instead, with knots at the sample points.

The paper of Good and Gaskins (1971) shows how one might prove that MPLE's are weakly consistent and also gives algorithms and some empirical material. In a much more recent paper, Good and Gaskins (1980) have followed up the 1971 paper and used the penalized likelihood methods in an exploratory data analysis mode as a method for bump-hunting. It is, of course, natural to expect that the use of penalized likelihood methods and penalized least squares splines are intimately related. In a very nice expository tract on this topic, Tapia and Thompson (1978) discuss the general problem of nonparametric penalized likelihood density estimation.

We close this section by noting the papers by Lii and Rosenblatt (1975) and Lii (1978). These papers focus on cubic spline interpolators of the empirical distribution function. They show in some cases that tail behaviors of the spline interpolators are not as well behaved as some more standard density estimators. This behavior is related to the specification of the boundary conditions.

9. SPLINES IN TIME SERIES ANALYSIS

Time series analysis is perhaps one of the richest disciplines from the point of view of having many functions to estimate nonparametrically. The most obvious of these functions is the spectral density which, of course, bears many analogies to the probability density. A distinctive feature of the spectral density is its periodic character with period 2π. This relatively inconspicous characteristic allows a somewhat more extensive spline theory to be developed.

Cogburn and Davis (1974) in another landmark paper develop the theory of periodic smoothing splines with application to spectral density estimation. They assume a model of the form

$$h(\omega) = f(\omega) + \epsilon(\omega), \quad \omega \in (0, 2\pi)$$

with

$$f \in W_m(0, 2\pi), \quad E\,\epsilon(\omega) = 0, \quad \omega \in (0, 2\pi)$$

and

$$E(\epsilon(\omega_1)\,\epsilon(\omega_2)) = \sigma^2 \quad \omega_1 = \omega_2$$
$$= 0 \quad \omega_1 \neq \omega_2$$

where h is observed either on a lattice of points or continuously and the noise variance σ^2 is unknown. The asymptotic solution devised by Cogburn and Davis is very convenient to handle, and easy to compute since it avoids explicit optimization.

In analogy with problem (3.1), Cogburn and Davis set out to solve the following problem:

$$\text{Minimize } \frac{1}{n}\sum_{j=1}^{2n}\left(f\left(\frac{k\pi}{n}\right) - h\left(\frac{k\pi}{n}\right)\right) + \lambda\int_0^{2\pi}(Lf(\omega))^2 d\omega$$

$$\text{subject to} \quad f \in W_m(0, 2\pi) \tag{9.1}$$

The solution s to this problem is called the *periodic lattice smoothing spline* (LSS). When h is known for all $\omega \in (0, 2\pi)$, problem (9.1) is replaced with

$$\text{Minimize} \quad \frac{1}{\pi}\int_0^{2\pi}(f(\omega) - h(\omega))^2 d\omega + \lambda\int_0^{2\pi}(L\,f(\omega))^2 d\omega$$

$$\text{subject to} \quad f \in W_m(0, 2\pi). \tag{9.2}$$

The solution to (9.2) is called the *periodic continuous smoothing spline* (CSS). Cogburn and Davis discuss algorithms for fitting the LSS and CSS. We summarize the final form.

Let $P(u)$ be the characteristic polynomial of L, say, $P(u) = u^m + \gamma_1 u^{m-1} + \cdots + \gamma_m$ and let $Q(k) = |P(ik)|^2$, $i = \sqrt{-1}$. Take $\lambda = 1/\alpha^{2m}$, $q_{n,j} = 1 + Q(j)\,[\sum_{l\neq 0} 1/Q(j + 2nl)]$, and let

$$a_{n,\alpha j} = \frac{1}{2n}\left(\frac{\alpha^{2m}}{Q(j) + \alpha^{2m}{}_{q_{n,j}}}\right),\ |j| \leq n$$

and $a_{n,\alpha l} = (Q(j)/Q(l))\,a_{n,\alpha j}$ for $l \in \{j \pm 2n, j \pm 4n, \ldots\}$, $|j| \leq n$. Then the LSS to h is given by the (discrete) convolution

$$h_*{}^n\, s_{n,\alpha}(\omega) = \sum_{j=-n+1} h\left(\frac{j\pi}{n}\right) s_{n,\alpha}\left(\omega - \frac{j\pi}{n}\right)$$

where $s_{n,\alpha}(\omega) = \sum_{k=-\infty}^{\infty} a_{n,\alpha,k}\, e^{ik\omega}$. Letting

$$a_{\alpha,l} = \lim_{n\to\infty} n\, a_{n,\alpha,l} = \frac{\lambda^{2m}}{2(Q(l) + \lambda^{2m})}$$

and $s_\alpha(\omega) = (1/\pi)\sum a_{\alpha,l}\, e^{ik\omega}$, we can closely approximate $n\, s_{n,\alpha}(\omega)$ by $\pi\, s_\alpha(\omega)$. The CSS to h is given by the convolution

$$h * s_\alpha(\omega) = \int_0^{2\pi} h(x)\, s_\alpha(\omega - x)dx. \tag{9.3}$$

If it is known that the function f to be estimated has derivatives of order m, but no specific operator L is known, a natural choice is $L = D^m$; in which case $s_\alpha(\omega)$ becomes

$$s_\alpha(\omega) = \frac{1}{2\pi}\sum_{l=-\infty}^{\infty}\frac{\lambda^{2m}}{l^{2m} + \alpha^{2m}}\, e^{il\omega}. \tag{9.4}$$

Writing

$$t_\alpha(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\alpha^{2m}}{\alpha^{2m} + y^{2m}}\, e^{i\omega y}\, dy$$

and

$$\hat{t}_\alpha(\omega) = \sum_{k=-\infty}^{\infty} t_\alpha(\omega + 2k\pi),$$

it is shown that $h * s_\alpha(\omega) = h * \hat{t}_\alpha(\omega) = \int_{-\infty}^{\infty} h(y)t_\alpha(\omega - y)dy$.

These results of Cogburn and Davis are particularly valuable because they are computationally easy and they facilitate further theoretical investigation by giving the LSS and CSS in a closed form. The remainder of Cogburn and Davis is devoted to using the periodic spline structure to estimate spectral densities.

Let $X_1, X_2, X_3, \ldots$ be a second-order stationary stochastic process with $EX_k = 0$ and $EX_jX_{j+k} = \sigma_k$. The σ_k are Fourier coefficients of a symmetric (spectral) distribution function, F, on $(-\pi, \pi)$. When F is absolutely continuous, it is completely determined by its spectral density

$$f(\omega) = DF(\omega) = \sum_{j=-\infty}^{\infty}\sigma_j e^{ij\omega}, \quad \omega \in (-\pi, \pi)$$

The statistical problem is to estimate $f(\omega)$ on the basis of a time series $X_1, \ldots, X_n$. The periodogram $I(\omega)$ is denoted by

$$I(\omega) = \sum_{k=-n+1}^{n-1}\hat{\sigma}_k e^{ik\omega}, \quad \omega \in (-\pi, \pi)$$

with

$$\hat{\sigma}_k = \frac{1}{n}\sum_{j=1}^{n-k} X_jX_{j+k}, \quad k = 0, 1, 2, \ldots, n - 1.$$

Since the periodogram is not a statistically consistent estimator of f, some modification is required. Smoothed estimators of $f(\omega)$ are obtained by smoothing the periodogram

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} I(x)K(\omega - x)dx \tag{9.5}$$

or by weighting the covariances by a *lag window* $k_m(j)$ giving

$$\hat{f}(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} k(j)\hat{\sigma}_j e^{ij\omega}, \quad \omega \in (-\pi, \pi) \tag{9.6}$$

where

$$K(\omega) = \frac{1}{2\pi}\sum_{j=-\infty}^{\infty} k(j)e^{ij\omega}, \quad \omega \in (-\pi, \pi).$$

The parallels of (9.3) with (9.5) and of (9.4) and (9.6) make clear the connection of periodic splines with kernel smoothers. In fact, (9.4) shows that the optimal lag window (measured against the criterion in (9.2) is, in fact,

$$k(j) = \frac{\alpha^{2m}}{j^{2m} + \alpha^{2m}} \quad j = 0, \pm 1, \pm 2 \ldots \tag{9.7}$$

where, of course, $\lambda = \alpha^{-2m}$ is the smoothing parameter. We note parenthetically that Parzen (1958) some 16 years prior to the Cogburn and Davis article had suggested the kernel in (9.7) as one having asymptotic efficiency of one. In a personal communication, Parzen indicated that he did not realize the connection with splines and had dismissed it then because of computational limitations.

The periodic smoothing splines, as we have seen, have an intimate connection with window estimators. Unfortunately, they do not fit well into the linear model, $h(\omega) = f(\omega) + \epsilon(\omega)$, as initially posited in this section. In fact, it is well known that

$$\frac{2I(\omega)}{f(\omega)} = \epsilon(\omega),$$

say, is asymptotically chi-squared with two degrees of freedom for $\omega \in (-\pi, \pi)$, $\omega \neq 0$. Thus a multiplicative model is more appropriate, or equivalently, a linear, additive model in logarithms

$$\log I(\omega) = \log f(\omega) + \epsilon^*(\omega)$$

The $\log f(\omega)$ is the so-called cepstrum. Wahba and Wold (1975b) and Wahba (1979c) discuss the use of periodic splines and cross-validation methods for estimating the cepstrum. Wahba, in particular, shows that an unbiased estimate of the expected integrated mean squared error can be obtained as a function of the smoothing parameter, λ. Results of Monte Carlo experiments are given as well.

In Wegman (1980a, 1981), the periodic splines of Cogburn and Davis are extended to the estimation of a vector-valued function. The results are applied not only to the estimation of spectral density matrices, but to a variety of other time series functions including phase, gain coherency, cepstrum, transfer function, and impulse response functions.

The estimation of time series functions was explored in a different direction by Peele and Kimeldorf (1977,1979) in a pair of papers we think are too little appreciated. If $T \subseteq I$ are sets of real numbers, then we let X_t, $t \in I$ be a real time series whose mean function is unknown but whose covariance kernel is assumed known. For each $i \in I$, X_i is predicted by a minimum mean squared error unbiased linear predictor $\hat{X}_i$ based on $\{X_t: t \in I\}$. If $\hat{x}_i$ is the evaluation of $\hat{X}_i$ based on a set of observations, the function $\hat{X}$ is called the prediction function. Mean estimation functions are defined by the authors in a similar way. For certain prediction and estimation problems, Peele and Kimeldorf characterized these functions in terms of the covariance structure of the process. In particular $\hat{X}$ is shown to be a spline function interpolating a convex set.

We believe the results of Peele and Kimeldorf are important ones. Box and Jenkins (1970) have widely popularized an approach to time series analysis based on the autoregressive-moving average (ARMA) model. For (mean) nonstationary time series, they find it useful to repeatedly difference the time series until apparent nonstationarities are removed.

The Peele-Kimeldorf procedure of fitting a spline is an interesting alternative to this differencing procedure.

The papers of Peele and Kimeldorf are extensions of a very important paper by Kimeldorf and Wahba (1970a). In an L-spline setting they take $B = [b_{jk}]$ to be a positive definitive matrix with inverse $B^{-1} = [b^{jk}]$.

Problem I: Find $f \in W_m(-\infty, \infty)$ which minimizes

$$\sum_{j,k} (f(t_j) - y_j)\, b^{jk}\, (f(t_k) - y_k) + \int_{-\infty}^{\infty} (Lf(t))^2 dt.$$

Problem II: Find f with $f(t) = E(X_t \mid Y_{t_1}, \ldots, Y_{t_n})$ where $Y_j = X_{tj} + \epsilon_j$ with $\epsilon_j \sim n(0, B)$ and X_t is a stationary Gaussian autoregressive process of order m.

Kimeldorf and Wahba show that the solution f of Problems I and II is the same function. They then go on to give a Bayesian interpretation.

A somewhat different time series/regression problem is addressed by Sacks and Ylvisaker (1966,1968,1970) and later by Eubank, Smith, and Smith (1981). They consider a stochastic process

$$Y_t = \beta f(t) + X_t, \quad t \in [0, 1]$$

where β is an unknown parameter, f is a known regression function, and X_t is a zero mean stochastic process with covariance function $\sigma(s, t)$. For finite sampling schemes, the regression design problem (cf Section 4) for estimating β has been addressed by Sacks and Ylvisaker (1966). They consider the problem of selecting a set of n distinct design points, $\{t_1, \ldots, t_n\}$, in the interval [0, 1] so that $\hat{\beta}(y_{t_1}, \ldots, y_{t_n})$ is the best linear unbiased estimator of β obtained by taking observations in the design set. For certain functions f and covariance functions $\sigma(s, t)$ they show the existence of optimal designs. There are difficulties constructing optimal designs and the authors are led to the construction of asymptotically optimal designs. Eubank, Smith, and Smith (1981) draw on some uniqueness results (see Barrow and Smith 1978), which show that the best L_2 approximation of a certain class of functions by piecewise polynomials with variable knots is unique. Eubank, Smith, and Smith show that solving the approximation problem is equivalent to finding the optimal design.

We note in closing this section that the literature in system theory and control has many parallels to time series analysis. Splines have become quite fashionable in that literature.

10. ISOTONE AND RELATED SPLINES

The theory of *isotonic regression*, also known as *statistical inference under order restrictions*, has received some attention over the years, but, we believe, has not fulfilled its full promise, partly because isotone estimators have disappointing continuity properties. However, because they share a fairly similar theoretical structure, the marriage of isotonic estimators with splines provides a nice vehicle for improving the continuity properties of isotonic estimates. A general exposition of the theory of isotonic regression may be found in Barlow et al. (1972).

To summarize briefly, we let $\preccurlyeq$ be a partial order (reflexive, transitive and antisymmetric) or a quasi-order (reflexive and transitive only) and let $\leq$ be the natural order on the real line. A function $f\colon R \to R$ is said to be *isotone* if it preserves the order; that is, if $x_1, x_2 \in R$ with $x_1 \preccurlyeq x_2$, then $f(x_1) \leq f(x_2)$. The general statistical problem is to estimate f from data in such a way that the estimator preserves the order; that is, the estimator is isotone. The set of isotone functions (call it W_I) is a closed, convex subset of L_2, and plays a role very analogous to W_m in spline theory. The penalized least squares spline is the (penalized) projection of the "data" onto W_m while the isotone estimator is the projection of the "data" onto W_I. The parallels between isotonic inference and spline theory are described at some length in Wegman (1980b).

The most frequently used examples of isotonicity include ordinary monotonicity and unimodality although, of course, there are many other examples as well. Wright and Wegman (1980) characterize isotonic functions by a continuous linear map, F, which maps W_m into L_2 and which commutes with D; that is, $D(Ff) = F(Df)$. A partial order, $\succcurlyeq$, on W_m is defined by $f \succcurlyeq 0$ if and only if $(Ff)(x) \geq 0$ for every $x \in [0, 1]$. The authors go on to show that a number of isotonic conditions on f can be characterized in this manner including nonnegativity, monotonicity, convexity, unimodality, and other compound order requirements. They then consider optimization problems of this sort:

$$\text{Minimize} \quad \int_0^1 (D^m f(x))^2 dx \quad \text{subject to}$$

$$\text{(i)} \quad f \in W_m$$

$$\text{(ii)} \quad f \succcurlyeq 0$$

$$\text{(iii)} \quad \alpha_i \leq f(x_i) \leq \beta_i, \quad i = 1, 2, \ldots, n. \qquad (10.1)$$

The main thrust of these authors is to show existence and a characterization of the solution to problem (10.1) as a $2m - 1$ degree polynomial spline. The latter paper goes on to give a statistical interpretation as a regression model, to give statistical consistency results and to make some remarks on computational algorithms. The recent paper of Laurent (1980) speaks in more detail to the problem of computation of restricted splines.

It is, in general, not too hard to believe that such isotonic splines should exist, since $W_m \cap W_I$ forms an ideal set over which to optimize. Computational algorithms are clearly the stumbling block in further development of the theory of isotone splines. When such algorithms become available we believe that smooth, order-preserving nonparametric estimators will substantially enhance the efficiency of estimation procedures currently in use.

The work of Wright and Wegman frames the isotonic spline problem in a general setting, but a number of papers attacking more limited problems in other settings have been published. Passow (1974) proves the existence of piecewise monotone spline interpolators. In a follow-up paper, Passow and Roulier (1977) focus on monotone and convex spline interpolation existence and characterization. All of these papers describe solutions to parts of the general optimization problem:

$$\text{Minimize} \quad \int_0^1 (D^m f(x))^2 dx \quad \text{subject to}$$

$$\text{(i)} \quad f \in W_m$$

$$\text{(ii)} \quad f \succcurlyeq 0$$

$$\text{(iii)} \quad f(x_i) = y_i, \quad i = 1, 2, \ldots, n.$$

In work more closely related to the regression splines of Section 4, De Vore (1977) describe monotone approximation by splines and Chui, Smith, and Ward (1980) describe techniques for the calculation of spline approximations to monotone nondecreasing functions using equally spaced knots.

11. GENERAL CONCLUDING REMARKS

We have attempted in this article to survey some of the uses of splines in statistics. In particular, splines may be regarded in some sense as optimal, nonparametric function estimators. The richness and diversity of points of view are evident by the extent of our list of references. Summarizing all of these references adequately is a task whose scope was not at first evident to us, but became clearer as we attempted to digest all of this material. No doubt in many respects we have failed, but we hope our perspective will stimulate interest and research in this very interesting and useful area.

There is no doubt that many interesting research problems abound. The computational issues, for example, are still largely undetermined. We have mentioned connections with isotonic methods and robust methods. The development of the theory of isotonic splines and of robust splines is still wide open. We mention the possibility of estimating functions in a reliability, a demographic, and a biomedical setting. The applications are completely untouched as far as we know. The applications to geology, geography, meteorology, and remote sensing at this stage are almost mere speculations made by a few knowledgeable persons. Even in the more well-exploited areas such as regression, there are many problems. The multidimensional surface estimation work, for example, has been developed largely in the last few years and has been largely the work of a handful of investigators. Generally speaking, statistical properties of any type of splines are not well developed. This is particularly true in terms of small-sample properties. Their use as smoothers in an exploratory data analysis context is, as far as we know, completely unexploited. We note also that McClure and Geman in a personal communication report that splines are a special case of nonparametric estimation by the method of sieves. (See Geman and Hwang 1982). They are currently exploring this new methodology, but it also is not yet developed. We believe the time is ripe for the study of splines and we hope this article will help crystallize thoughts in this direction.

[*Received June 1981. Revised August 1982.*]

REFERENCES

AGARWAL, G.G., and STUDDEN, W.J. (1978), "Asymptotic Design and Estimation Using Linear Splines," *Communications in Statistics: Simulation and Computation B*, 7, 309–320.

AHLBERG, J.H., NILSON, E.N., and WALSH, J.L. (1967), *The Theory of Splines and Their Applications*, New York: Academic Press.

AMOS, D., and SLATER, M. (1969), "Polynomial and Spline Approximation by Quadratic Programming," *Communications of the Association for Computing Machinery*, 12, 379ff.

ANDERSSEN, R., BLOOMFIELD, P., and McNEIL, D. (1974), "Spline Functions in Data Analysis," Technical Report 69, Princeton University, Department of Statistics.

ANSELONE, P.M., and LAURENT, P.J. (1968), "A General Method for the Construction of Interpolating or Smoothing Spline Functions," *Numerische Mathematik*, 12, 66–82.

ARONSZAJN, N. (1950), "Theory of Reproducing Kernels," *Transactions of the American Mathematical Society*, 68, 337–404.

ATTEIA, M. (1968), "Fonctions (Spline) Definies Sur un Ensemble Convexe," *Numerische Mathematik*, 12, 192–210.

——— (1970), "Fonctions Spline et Noyaux Reproduisants d'Aronszajn et Bergman," *Revue Française d'Informatique et de Recherche Operationelle*, R3, 31–43.

BARLOW, R.E., BARTHOLOMEW, D.J., BREMNER, J.M., and BRUNK, H.D. (1972), *Statistical Inference Under Order Restrictions*, New York: John Wiley.

BARROW, D.L., and SMITH, P.W. (1978), "Asymptotic Properties of Best $L_2(0, 1)$ Approximation by Splines with Variable Knots," *Quarterly of Applied Mathematics*, 36, 293–304.

BONEVA, L., KENDALL, D., and STEFANOV, I. (1971), "Spline Transformations: Three New Diagnostic Aids for the Data Analyst," *Journal of the Royal Statistical Society*, Ser. B, 33, 1–70.

BOX, G.E.P., and JENKINS, G.M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden-Day.

BUSE, A., and LIM, L. (1977), "Cubic Splines as a Special Case of Restricted Least Squares," *Journal of the American Statistical Association*, 72, 64–68.

CHUI, C.K., SMITH, P.W., and WARD, D.J. (1980), "Degree of L_p Approximation by Monotone Splines," *SIAM Journal of Mathematical Analysis*, 11, 436–447.

CLARK, R.M. (1977), "Nonparametric Estimation of Smooth Regression Function," *Journal of the Royal Statistical Society*, Ser. B, 39, 107–113.

COGBURN, R., and DAVIS, H.T. (1974), "Periodic Splines and Spectral Estimation," *Annals of Statistics*, 2, 1108–1126.

COPLEY, P., and SCHUMAKER, L.L. (1978), "On pLg-Splines," *Journal of Approximation Theory*, 23, 1–28.

CRAVEN, P., and WAHBA, G. (1979), "Smoothing Noisy Data With Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerische Mathematick*, 31, 377–403.

DANIEL, J.W., and SCHUMAKER, L.L. (1974), "On the Closedness of the Linear Image of a Set With Application to Generalized Spline Functions," *Applicable Mathematics*, 4, 191–205.

DELFINER, P. (1978), "The Intrinsic Model of Order *k*," Ecole des Mines De Paris, Summer School Notes.

DEREK, J. (1966), "Fitted Segmented Curves Whose Joint Points Have to be Estimated," *Journal of the American Statistical Association*, 61, 1097–1129.

DE VORE, R. (1977), "Monotone Approximation by Splines," *SIAM Journal of Mathematical Analysis*, 8, 891–905.

DRAPER, N.R., GUTTMAN, I., and LIPOW, P. (1977), "All-Bias Designs for Splines Joined at the Axes," *Journal of the American Statistical Association*, 72, 424–429.

DUCHON, J. (1976a), "Interpolation des Fonctions de Deux Variables Suivant le Principe de la Flexion des Plaques Minces," *R.A.I.R.O. Analyse Numerique*, 10, 12, 5–12.

——— (1976b), "Fonctions Spline du Type 'Plaque Mince' en Dimension 2," No. 231, Seminaire d'analyse Numerique, Mathematiques Appliquees, University Scientifique et Medicale de Grenoble.

ERTEL, J.E., and FOWLKES, E.B. (1976), "Some Algorithms for Linear Spline and Piecewise Multiple Linear Regression," *Journal of the American Statistical Association*, 71, 640–648.

EUBANK, R., SMITH, P.L., and SMITH, P.W. (1981), "Uniqueness and Eventual Uniqueness of Optimal Designs in Some Time Series Models," *Annals of Statistics*, 9, 486–493.

FLEISHER, J. (1979), "Spline Smoothing Routines," Reference Manual for the 1110, Academic Computer Center, The University of Wisconsin, Madison.

FRIEDMAN, J., GROSSE, E., and STUETZLE, W. (1980), "Multidimensional Additive Spline Approximation," SLAC PUB-2504, Stanford University.

FRIEDMAN, J., and STUETZLE, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823.

FULLER, WAYNE A. (1969), "Grafted Polynomials as Approximating Functions," *Australian Journal of Agricultural Economics*, 13, 35–46.

GALLANT, R., and FULLER, W. (1973), "Fitting Segmented Polynomial Regression Models Whose Join Points have to be Estimated," *Journal of the American Statistical Association*, 68, 144–147.

GAMBER, H. (1979a), "Choice of an Optimal Shape Parameter When Smoothing Noisy Data," *Communications in Statistics: Theory and Methods*, 14, 1425–1435.

——— (1979b), "Confidence Regions for Periodic Functions," *Communications in Statistics: Theory and Methods*, 14, 1437–1446.

GEMAN, S., and HWANG, C.R., (1982), "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *Annals of Statistics*, 10, 401–414.

GOLUB, G., HEATH, M., and WAHBA, G. (1979), "Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter," *Technometrics*, 21, 215–224.

GOOD, I.J., and GASKINS, R.A. (1971), "Non-Parametric Roughness Penalties for Probability Densities," *Biometrika*, 58, 255–277.

——— (1980), "Density Estimation and Bump-Hunting by Penalized Likelihood Method Exemplified by Scattering and Meteorite Data" (invited paper with discussion), *Journal of the American Statistics Association*, 75, 42–73.

GREVILLE, T.N.E. (ed.) (1969), *Theory and Application of Spline Functions*, New York: Academic Press.

GUARASCIO, M., DAVID, M., and HUIJBREGTS, C. (eds.) (1976), *Advanced Geostatistics in the Mining Industry*, D. Reidel, Dordrecht-Holland.

HUBER, P.J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.

KARLIN, S., MICHELLI, C.A., PINKUS, A., and SCHOENBERG, I.J. (1976), *Studies in Spline Functions and Approximation Theory*, New York: Academic Press.

KIMELDORF, G.S., and WAHBA, G. (1970a), "A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines," *Annals of Mathematical Statistics*, 41, 495–502.

——— (1970b), "Spline Functions and Stochastic Process," *Sankhya (A)*, 32, 173–180.

——— (1971), "Some Results on Tchebycheffian Spline Functions," *Journal of Mathematical Analysis and Application*, 33, 82–94.

LAURENT, P.J. (1969), "Construction of Spline Functions in a Convex Set," *Approximation with Special Emphasis on Spline Functions* (ed. I.J. Schoenberg), New York: Academic Press, 415–446.

——— (1972), *Approximation et Optimisation*, Paris: Hermann.

——— (1980), "An Algorithm for the Computation of Spline Functions with Inequality Constraints," Laboratorie de Mathematiques Appliquees, Universite de Grenoble.

LENTH, R.V., (1977), "Robust Splines," *Communications in Statistics: Theory and Methods*, 6, 847–854.

LII, K.S. (1978), "A Global Measure of a Spline Density Estimate," *Annals of Statistics*, 6, 1138–1148.

LII, K.S., and ROSENBLATT, M. (1975), "Asymptotic Results on a Spline Estimate of a Probability Density," *Stochastic Processes and Their Applications*, 2, 77–86.

LYCHE, T., and SCHUMAKER, L.L. (1973), "Computation of Smoothing and Interpolating Spline via Local Basis," *SIAM Journal Numerical Analysis*, 10, 1027–1038.

MANGASARIAN, O.L., and SCHUMAKER, L.L. (1969), "Spline via Optimal Control," *Approximations with Special Emphasis on Spline Functions* (ed. I.J. Schoenberg), New York: Academic Press, 119–156.

MATHERON, G. (1973), "The Intrinsic Random Functions and Their Applications," *Advances in Applied Probability*, 5, 439–468.

MEINGUET, J. (1978), "Multivariate Interpolation at Arbitrary Points Made Simple," Report No. 118, Institute de Mathematique Pure et Appliquee, Universite Catholique de Louvain, to appear ZAMD.

——— (1979), "An Intrinsic Approach to Multivariate Spline Interpolation at Arbitrary Points," *Proceedings of the Nato Advanced Study Institute on Polynomial and Spline Approximation*, Calgary, 1978.

MERZ, P. (1978), "Spline Smoothing by Generalized Cross-Validation, a Technique for Data Smoothing," Chevron Research Company, Richmond, CA.

DEMONTRICHER, G.F., TAPIA, R.A., and THOMPSON, J.R. (1975), "Non-Parametric Maximum Likelihood Estimation of Probability Densities by Penalty Function Methods," *Annals of Statistics*, 3, 1329–1348.

MURTY, V.N. (1971), "Optimal Designs with a Single Multiple Knot at the Center," *Annals of Mathematical Statistics*, 42, 952–960.

PAIHUA-MONTES, L. (1979), *Quelques Methodes Numeriques Pour le Calcul de Fonctions Splines a Une et Plusiers Variable*, Thesis, Universite Scientifique et Medicale de Grenoble.

PARK, S.H. (1978), "Experimental Designs for Fitting Segmented Polynomial Regression Models," *Technometrics*, 20, 151–154.

PARZEN, E. (1958), "On Asymptotically Efficient Consistent Estimates of the Spectral Density Function of a Time Series," *Journal of the Royal Statistical Society*, Ser. B, 20, 303–322.

——— (1961), "An Approach to Time Series Analysis," *Annals of Mathematical Statistics*, 32, 951–989.

PASSOW, E. (1974), "Piecewise Monotone Spline Interpolation," *Journal of Approximation Theory*, 12, 240–241.

PASSOW, E., and ROULIER, J.A. (1977), "Monotone and Convex Spline Interpolation," *SIAM Journal Numerical Analysis*, 14, 904–909.

PEELE, L., and KIMELDORF, G. (1977), "Prediction Functions and Mean Estimation Functions for a Time Series," *Annals of Statistics*, 5, 709–721.

——— (1979), "Time Series Prediction Functions Based on Imprecise Observations," *Annals of Statistics*, 7, 801–811.

POIRER, D.J. (1973), "Piecewise Regression Using Cubic Splines," *Journal of the American Statistical Association*, 68, 515–524.

PRENTER, P.M. (1975), *Splines and Variational Methods*, New York: John Wiley.

REINSCH, C.H. (1967), "Smoothing by Spline Functions I," *Numerische Mathematik*, 10, 177–183.

——— (1971), "Smoothing by Spline Function II," *Numerische Mathematik*, 16, 451–454.

RENDU, J.-M. (1978), *An Introduction to Geostatistical Methods of Mineral Evaluation*, South African Institute of Mining and Metallurgy, Johannesberg.

RITTER, K. (1969), "Generalized Spline Interpolation and Non-Linear Programming," *Approximation with Special Emphasis on Spline Functions*, (ed. I.J. Schoenberg), New York: Academic Press, 75–118.

ROSENBLATT, M. (1971), "Curve Estimates," *Annals of Mathematical Statistics*, 42, 1815–1842.

SACKS, J., and YLVISAKER, D. (1966), "Designs for Regression Problems with Correlated Errors," *Annals of Mathematical Statistics*, 37, 66–89.

——— (1968), "Designs for Regression Problems with Correlated Errors: Many Parameters," *Annals of Mathematical Statistics*, 39, 49–69.

——— (1970), "Designs for Regression Problems With Correlated Errors: III," *Annals of Mathematical Statistics*, 41, 2057–2074.

SCHOENBERG, I.J. (1964), "Spline Functions and the Problem of Graduation," *Proceedings of the National Academy of Sciences of the U.S.A.*, 52, 947–950.

SCHOENBERG, I.J. (ed.) (1969), *Approximations with Special Emphasis on Spline Functions*, New York: Academic Press.

SCHULTZ, M., and VARGA, R. (1967), "*L*-Splines," *Numerische Mathematik*, 10, 345–369.

SMITH, PATRICIA (1979), "Splines as a Useful and Convenient Statistical Tool," *The American Statistician*, 33, 57–62.

SPECKMAN, P. (1983), "Splines Smoothing and Optimal Rates of Convergence in Nonparametric Regression Models," to appear *Annals of Statistics*

STONE, C. (1980), "Optimal Rates of Convergence for Nonparametric Estimators," *Annals of Statistics*, 8, 1348–1360.

——— (1982), "Optimal Global Rates of Convergence for Nonparametric Regression," *Annals of Statistics*, 10, 1040–1053.

STUDDEN, W., and VAN ARMAN, D. (1969), "Admissible Designs for Polynomial Spline Regression," *Annals of Mathematical Statistics*, 40, 1557–1569.

TAPIA, R.A., and THOMPSON, J.R. (1978), *Nonparametric Probability Density Estimation*, Baltimore: Johns Hopkins University Press.

UTRERAS-DIAS, F. (1979), "Cross-Validation Techniques for Smoothing Spline Functions in One or Two Dimensions," *Smoothing Techniques for Curve Estimation* (eds. T. Gasser and M. Rosenblatt), Berlin: Springer-Verlag, 196–232.

WAHBA, G. (1971), "A Polynomial Algorithm for Density Estimation," *Annals of Mathematical Statistics*, 42, 1870–1886.

——— (1973), "On the Minimization of a Quadratic Functional Subject to a Continuous Family of Linear Inequality Constraints," *SIAM Journal of Control*, 11, 64–79.

——— (1975a), "Optimal Convergence Properties of Variable Knot, Kernel and Orthogonal Series Estimates for Density Estimation," *Annals of Statistics*, 3, 15–29.

——— (1975b), "Interpolating Spline Methods for Density Estimation I, Equi-Spaced Knots," *Annals of Statistics*, 3, 30–48.

——— (1975c), "Smoothing Noisy Data by Spline Functions," *Numerische Mathematik*, 24, 383–393.

——— (1976), "A Survey of Some Smoothing Problems and the Method of Generalized Cross-Validation for Solving Them," *Applications of Statistics* (ed. P.R. Krishnaiah), North Holland, 507–524.

——— (1977), "Practical Approximate Solutions to Linear Operator Equations When the Data are Noisy," *SIAM Journal Numerical Analysis*, 14, 651–667.

——— (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society*, Ser. B, 40, 364–372.

——— (1979a), "Convergence Rates of 'Thin Plate' Smoothing Splines When the Data are Noisy," *Smoothing Techniques for Curve Estimation*, (eds. T. Gasser and M. Rosenblatt), Berlin: Springer-Verlag, 233–245.

——— (1979b), "How to Smooth Curves and Surfaces with Splines and Cross-Validation," *Proceedings of the 24th Conference on the Design of Experiments*, U.S. Army Research Office, Report 79-2.

——— (1979c), "Automatic Smoothing of the Log Periodogram," *Journal of the American Statistical Association*, 75, 122–132.

——— (1980a), "On Mildly, Moderately and Severely Ill-posed Problems," *Proceedings of the International Symposium Ill-posed Problems* (ed. M.Z. Nashed).

——— (1980b), "Spline Bases, Regularization, and Generalized Cross-validation for Solving Approximation Problems with Large Quantities of Noisy Data," *Proceedings of the International Conference on Approximation Theory in Honor of George Lorenz*, Jan. 8–10, Austin, TX, (ed. Ward Cheney), Academic Press.

——— (1981), "Spline Interpolation and Smoothing on the Sphere," *SIAM Journal of Scientific and Statistical Computing*, 2, 5–16.

WAHBA, G., and DYN, N. (1982), "On the Estimation of Functions of Several Variables from Aggregated Data," *SIAM Journal of Mathematical Analysis*, 13, 134–152.

WAHBA, G., and MICCHELI, C. (1981), "Problems for Optimal Surface Interpolation," *Approximation Theory and Applications*, ed. Z. Ziegler, New York: Academic Press.

WAHBA, G., and WENDELBERGER, J. (1980), "Some New Mathematical Methods for Variational Object Analysis Using Splines and Cross-validation," *Monthly Weather Review*, 108, 36–57.

WAHBA, G., and WOLD, S. (1975a), "A Completely Automatic French Curve: Fitting Spline Functions by Cross-Validation," *Communications in Statistics*, 4, 1–17.

——— (1975b), "Periodic Splines for Spectral Density Estimation: The Use of Cross-validation for Determining the Degree of Smoothing," *Communications in Statistics*, 4, 125–141.

WEGMAN, E.J. (1972a), "Nonparametric Probability Density Estimation, I. A. Summary of Available Methods," *Technometrics*, 14, 533–546.

——— (1972b), "Nonparametric Probability Density Estimation, II. A Comparison of Density Estimation Methods," *Journal of Statistical Computation and Simulation*, 1, 225–245.

——— (1980a), "Optimal Estimation of Time Series Functions," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-28, 763–767.

——— (1980b), "Two Approaches to Nonparametric Regression: Splines and Isotonic Inference," *Recent Developments in Statistical*

Inference and Data Analysis (ed. K. Matusita), North-Holland, 323–334.
——— (1981), "Vector Splines and the Estimation of Filter Functions," *Technometrics*, 23, 83–89.
WEINERT, H.L. (1980), "Statistical Methods in Optimal Curve Fitting," *Advanced Concepts in Ocean Measurements for Marine Biology* (eds. F.P. Diemer, F.J. Vernberg, D.Z. Mirkes), Columbia, S.C.: U. of South Carolina Press.
WOLD, S. (1974), "Spline Functions in Data Analysis," *Technometrics*, 16, 1–11.
WRIGHT, I.W., and WEGMAN, E.J. (1980), "Isotonic, Convex and Related Splines," *Annals of Statistics*, 8, 1023–1035.