The Annals of Statistics
1979, Vol. 7, No. 4, 801–811

TIME SERIES PREDICTION FUNCTIONS BASED ON IMPRECISE OBSERVATIONS

BY LAWRENCE PEELE[1] AND GEORGE KIMELDORF[2]

Old Dominion University and University of Texas at Dallas

Let $T \subseteq I$ be sets of real numbers. Let $\{Y(t) : t \in I\}$ be a real time series whose covariance kernel is assumed known and positive definite. The mean is assumed either to be known or to be an unknown member of a known class of functions on I. For each fixed $s \in I$, $Y(s)$ is predicted by a minimum mean square error unbiased linear predictor $\hat{Y}(s)$ based on $\{Y(t) : t \in T\}$. If $\hat{y}(s)$ is the evaluation of $\hat{Y}(s)$ given that the sample path for $\{Y(t) : t \in T\}$ is an unknown element of a known collection of functions on T, then $\hat{y}(s)$ is a prediction for $Y(s)$ and the function $\hat{y}$ is called a prediction function. Mean-estimation functions are defined similarly. For certain prediction problems based on imprecise observations, characterizations are obtained for these functions in terms of the covariance structure of the process. For a particular prediction problem $\hat{y}$ is shown to be a spline function interpolating a convex set.

1. Introduction. Let I denote a set of real numbers and let $\{Y(t) : t \in I\}$ be a real time series of the form

$$Y(t) = m_0(t) + X(t), \tag{1.1}$$

where $\{X(t)\}$ has mean 0 and known positive definite covariance kernel k given by $k(s, t) = E[X(s)X(t)]$. The mean function m_0 for $\{Y(t)\}$ is assumed either to be known or to be an unknown member of a known class M of functions on I. Given any subset T of I, let $L[Y(t) : t \in T]$ denote the vector space of finite linear combinations of elements of $\{Y(t) : t \in T\}$ with inner product given by $\langle U, V\rangle = \mathrm{Cov}(U, V)$. Denote the completion of this inner-product space by $L^2[Y(t) : t \in T]$, so that $L^2[Y(t) : t \in T]$ is the Hilbert space generated by $\{Y(t) : t \in T\}$ with inner product determined by $\langle Y(s), Y(t)\rangle = k(s, t)$.

For each $s \in I$ let $Y(s)$ be predicted by an element $\hat{Y}(s) \in L^2[Y(t) : t \in T]$. Suppose that the sample path for $\{Y(t) : t \in T\}$ is not known, but it is known that the sample path belongs to a certain known collection of functions on T. An evaluation $\hat{y}(s)$ for $\hat{Y}(s)$ is made by choosing a statistically likely sample path for $\{Y(t) : t \in T\}$ and then evaluating $\hat{Y}(s)$ based on the chosen sample path. Thus, $\hat{y}(s)$ is a prediction for $Y(s)$, and the function $\hat{y}$ is called a *prediction function*.

Received January 1977; revised July 1978.

[1]Research conducted in part at Florida State University and supported by the United States Army Research Office, Durham, under Grant No. DAHC04-74-G-0188.

[2]Research conducted in part at Florida State University and supported by the Air Force Office of Scientific Research under Grant AFOSR-74-2581A.

AMS 1970 *subject classifications.* Primary 62M10, 41A15.

Key words and phrases. Prediction functions, mean-estimation functions, imprecise observations, spline functions, time series.

If the mean function m_0 is unknown, then for each $s \in I$, $m_0(s)$ is estimated by an element $\hat{Z}(s) \in L^2[Y(t) : t \in T]$. An evaluation $\hat{z}(s)$ for $\hat{Z}(s)$ is made, and the function $\hat{z}$ is called a *mean-estimation function.*

Consider the case when $T = \{t_1, t_2, \cdots, t_n\}$. Let $\mathbf{Y} = [Y(t_1), Y(t_2), \cdots, Y(t_n)]'$. Let $S \subseteq R^n$ and suppose that S is observed in the sense that it is observed that $\mathbf{Y} \in S$. So, although the actual value $\boldsymbol{\lambda}^0$ of $\mathbf{Y}$ is not observed, it is observed that $\mathbf{Y} \in S$. For example, rather than observing $Y(t_j)$ directly, we might be able to observe only the greatest integer that is less than or equal to $Y(t_j)$. On the other hand, if S consists of a single n-vector, then this type of observation reduces to the usual precise observation. Let $\hat{\boldsymbol{\lambda}} = [\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_n]'$ be an estimate for $\boldsymbol{\lambda}^0$. If, for each $s \in I$, $Y(s)$ is predicted by a random variable $\Sigma_{j=1}^n \hat{c}_j(s) Y(t_j)$, then the function $\hat{y}$ defined by

$$\hat{y}(s) = \Sigma_{j=1}^n \hat{c}_j(s)\hat{\lambda}_j \quad \text{for} \quad s \in I \tag{1.2}$$

is a prediction function. Similarly if, for each $s \in I$, $m_0(s)$ is estimated by a random variable $\Sigma_{j=1}^n \hat{d}_j(s) Y(t_j)$, then the function $\hat{z}$ defined by

$$\hat{z}(s) = \Sigma_{j=1}^n \hat{d}_j(s)\hat{\lambda}_j \quad \text{for} \quad s \in I \tag{1.3}$$

is called a mean-estimation function.

In this paper prediction functions and mean-estimation functions based on imprecise observations are characterized in terms of the reproducing kernel Hilbert space (RKHS) with reproducing kernel k. Certain prediction functions based on imprecise observations are shown to be types of spline functions; for example, one prediction function is a spline function interpolating a convex set.

2. Preliminaries and notation. For fixed $s \in I$, a random variable $\hat{Y}(s)$ is called a *minimum mean square error unbiased linear* (MEUL) *predictor* for $Y(s)$ if among random variables W satisfying the conditions

$$\text{unbiasedness:} \quad E_m(W) = m(s) \quad \text{for all} \quad m \in M, \tag{2.1}$$

$$\text{linearity:} \quad W \in L^2[Y(t) : t \in T], \tag{2.2}$$

the minimum of $E[W - Y(s)]^2$ occurs when $W = \hat{Y}(s)$. Similarly, a random variable $\hat{Z}(s)$ is called a *minimum variance unbiased linear* (MVUL) estimator for $m_0(s)$ if $\hat{Z}(s)$ is a random variable W which has minimal variance among random variables W satisfying (2.1) and (2.2).

Let I be an interval on the real line and $\mathfrak{L}$ be a qth order linear differential operator of the form

$$\mathfrak{L} = \Sigma_{i=0}^q a_i \mathfrak{D}^i, \tag{2.3}$$

where the functions $a_0, a_1, \cdots, a_q$ on I have q continuous derivatives. A function $\hat{g}$ is called an *$\mathfrak{L}$-spline of interpolation* to a set of points $\{(t_1, \lambda_1), (t_2, \lambda_2), \cdots, (t_n, \lambda_n)\}$ if, among functions g in a certain class (depending on $\mathfrak{L}$) of functions on I, $\hat{g}$ minimizes

$$\int_I [(\mathfrak{L}g)(t)]^2 \, dt$$

subject to the constraints

$$g(t_j) = \lambda_j \qquad \text{for} \quad j = 1, 2, \cdots, n.$$

Let I be a set of real numbers. Each nonnegative definite kernel k^* on $I \times I$ is the reproducing kernel for a unique RKHS which will be denoted $H(k^*)$. Let $k^*(\cdot, t)$ denote the element of $H(k^*)$ given by $[k^*(\cdot, t)](s) = k^*(s, t)$ for $s \in I$. For $T \subseteq I$, let $L[k^*(\cdot, t) : t \in T]$ denote the collection of finite linear combinations of elements of $\{k^*(\cdot, t) : t \in T\}$, and let $L^2[k^*(\cdot, t) : t \in T]$ denote the closure of $L[k^*(\cdot, t) : t \in T]$ in $H(k^*)$.

3. Background material. Previous papers on prediction functions have assumed that observations are precise. Kimeldorf and Wahba (1970a) showed that for a particular type of Gaussian time series with known mean, a prediction function, with prediction based on conditional expectation, is a particular type of $\mathfrak{L}$ spline. For a particular time series prediction problem with unknown mean, Kimeldorf and Wahba (1970b) showed that an MEUL prediction function is an $\mathfrak{L}$ spline. Peele and Kimeldorf (1977) extended the results of Kimeldorf and Wahba (1970b) and characterized mean-estimation functions for certain estimation problems. The following generalization by Peele and Kimeldorf (1977) of a lemma of Kimeldorf and Wahba (1970b) will be used in the present paper.

LEMMA 3.1. *Let $H = H_1 \oplus H_2$ be the direct orthogonal sum of real Hilbert spaces H_1 and H_2. Let P_i be the projection operator onto H_i and J be a closed subspace of H such that $P_1(J) = H_1$. Then for any given elements, $\tilde{w}, \tilde{u} \in H$:*

(a) *there exists a unique element $w = \hat{w} \in J$ which minimizes $\|\tilde{w} - w\|^2$ subject to the constraint $P_1(\tilde{w} - w) = 0$;*

(b) *there exists a unique element $u = \hat{u} \in H$ for which $\|P_2(u)\|^2$ is minimized among elements u satisfying $\tilde{u} - u \in J^{\perp}$;*

(c) $\langle \tilde{u}, \hat{w} \rangle = \langle \hat{u}, \tilde{w} \rangle$.

4. Prediction with known mean. Suppose that $\{Y(t) : t \in I\}$ has the model (1.1) with $m_0(t) \equiv 0$; that is, M consists only of the zero function. Suppose for some $S \subseteq R^n$ it is observed that $\mathbf{Y} \in S$. That is, the n-vector value $\boldsymbol{\lambda}^0$ taken on by the mean-zero n-vector $\mathbf{Y}$ is an unknown member of the known set S. Let $K = [k(t_i, t_j)]_{n \times n}$ and suppose that there exists an element $\hat{\boldsymbol{\lambda}} \in S$ minimizing

$$\boldsymbol{\lambda}' K^{-1} \boldsymbol{\lambda} \tag{4.1}$$

for $\boldsymbol{\lambda} \in S$. Statistical justification for estimating $\boldsymbol{\lambda}^0$ by an element $\hat{\boldsymbol{\lambda}} \in S$ which minimizes (4.1) for $\boldsymbol{\lambda} \in S$ is provided by Lemma 4.1 and Lemma 4.2.

The following result is well known.

LEMMA 4.1. *If $\mathbf{Y}$ is multivariate normal with mean zero and covariance matrix K, then an n-vector $\hat{\boldsymbol{\lambda}}$ minimizes* (4.1) *for $\boldsymbol{\lambda} \in S$ if and only if $\boldsymbol{\lambda}$ maximizes the likelihood function for $\boldsymbol{\lambda} \in S$.*

Now suppose that $\mathbf{Z} = [Z_1, Z_2, \cdots, Z_n]'$ is an n-vector of orthonormal, mean-zero random variables and $\mathbf{Z}$ is not necessarily multivariate normal. If $\mathbf{Z}$ is known to have taken on some unknown vector value $\boldsymbol{\beta}^0$ in a known subset S of R^n, then an intuitively appealing estimate for $\boldsymbol{\beta}^0$ is any n-vector $\hat{\boldsymbol{\beta}} \in S$ minimizing $E(\mathbf{Z} - \boldsymbol{\beta})'(\mathbf{Z} - \boldsymbol{\beta}) = \Sigma_{j=1}^n E(Z_j - \beta_j)^2$ for $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]' \in S$. Equivalently, $\boldsymbol{\beta}^0$ is estimated by an n-vector $\hat{\boldsymbol{\beta}} \in S$ minimizing $\Sigma_{j=1}^n \beta_j^2$ for $\boldsymbol{\beta} \in S$.

Let $\mathfrak{T}$ be a one-to-one linear operator from R^n onto R^n. Observing that $\mathbf{Y} \in S$ is equivalent to observing that the random vector $\mathfrak{T}(\mathbf{Y}) \in \mathfrak{T}(S)$.

LEMMA 4.2. *An n-vector $\hat{\boldsymbol{\lambda}}$ minimizes* (4.1) *for $\boldsymbol{\lambda} \in S$ if and only if for each one-to-one linear operator $\mathfrak{T}$ on R^n such that $\mathfrak{T}(\mathbf{Y})$ is an n-vector of orthonormal random variables, $\hat{\boldsymbol{\lambda}}$ minimizes $E(\mathfrak{T}(\mathbf{Y}) - \mathfrak{T}(\boldsymbol{\lambda}))'(\mathfrak{T}(\mathbf{Y}) - \mathfrak{T}(\boldsymbol{\lambda}))$ for $\boldsymbol{\lambda} \in S$.*

PROOF. Each one-to-one linear operator $\mathfrak{T}$ from R^n to R^n corresponds to a unique $n \times n$ nonsingular matrix $G_{\mathfrak{T}}$ in that $\mathfrak{T}(\boldsymbol{\lambda}) = G_{\mathfrak{T}}\boldsymbol{\lambda}$ for all $\boldsymbol{\lambda} \in R^n$. It can be easily seen that, for any nonsingular matrix G, the covariance matrix for the n-vector $G\mathbf{Y}$ of random variables is GKG'. Hence, the random variables in the n-vector $G\mathbf{Y}$ are orthonormal if and only if $GKG' = I$, or, equivalently, $G'G = K^{-1}$. Hence, if $\mathfrak{T}$ is a one-to-one operator from R^n to R^n, then $\mathfrak{T}(\mathbf{Y})$ is an n-vector of orthonormal random variables if and only if the matrix $G_{\mathfrak{T}}$ satisfies $G_{\mathfrak{T}}'G_{\mathfrak{T}} = K^{-1}$. Hence, if $\mathfrak{T}(\mathbf{Y})$ is an n-vector of orthonormal random variables, then it follows that for $\boldsymbol{\lambda} \in R^n$,

$$\begin{aligned}(\mathfrak{T}(\boldsymbol{\lambda}))'(\mathfrak{T}(\boldsymbol{\lambda})) &= (G_{\mathfrak{T}}\boldsymbol{\lambda})'(G_{\mathfrak{T}}\boldsymbol{\lambda}) \\ &= \boldsymbol{\lambda}'G_{\mathfrak{T}}'G_{\mathfrak{T}}\boldsymbol{\lambda} \\ &= \boldsymbol{\lambda}'K^{-1}\boldsymbol{\lambda}.\end{aligned}$$

Thus, in view of Lemma 4.1 and Lemma 4.2, the unknown vector $\boldsymbol{\lambda}^0$ will be estimated by a vector $\hat{\boldsymbol{\lambda}}$ which minimizes (4.1) for $\boldsymbol{\lambda} \in S$ if such a minimizing vector exists.

THEOREM 4.1. *Let $\{Y(t) : t \in I\}$ have the model* (1.1) *with $m_0 \equiv 0$. For fixed $t \in I$, let $\Sigma_{j=1}^n \hat{c}_j(t)Y(t_j)$ be a random variable W minimizing $E(W - Y(t))^2$ for $W \in L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$. Suppose that for some $S \subseteq R^n$, it is observed that $\mathbf{Y} \in S$. If there exists $\hat{\boldsymbol{\lambda}} \in S$ which minimizes* (4.1) *for $\boldsymbol{\lambda} \in S$, then the prediction function $\hat{y}$, given by* (1.2), *minimizes*

$$\|y\|_{H(k)}^2$$

among all functions satisfying

(a) $y \in H(k)$

and

(b) $[y(t_1), y(t_2), \cdots, y(t_n)]' \in S$.

PROOF. Let $J = L[k(\cdot, t_1), k(\cdot, t_2), \cdots, k(\cdot, t_n)]$. Let $H_1 = \{0\}$ and $H = H_2 = H(k)$. Note that $H(k)$ is isometric to $L^2[Y(t) : t \in I]$ and J is isometric to $L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$ under the mapping taking $k(\cdot, t)$ to $Y(t)$. Now apply Lemma 3.1 as follows. Let $\boldsymbol{\lambda} \in R^n$. If $\mathbf{c} = K^{-1}\boldsymbol{\lambda}$, then $u_{\boldsymbol{\lambda}} = \Sigma_{j=1}^n c_j k(\cdot, t_j)$ satisfies $u_{\boldsymbol{\lambda}}(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$. It can be easily seen that $u_{\boldsymbol{\lambda}}$ is the unique function of minimal norm among those functions $u \in H(k)$ satisfying $u(t_j) = \lambda_j$ for $j =$

$1, 2, \cdots, n$. Furthermore, $\|u_\lambda\|^2 = (K^{-1}\lambda)'K(K^{-1}\lambda) = \lambda'K^{-1}\lambda$. Hence, it remains only to show that $\hat{y} = u_{\hat{\lambda}}$. For fixed $t \in I$, $\Sigma_{j=1}^n \hat{c}_j(t)Y(t_j)$ is the unique random variable $P_Q(Y(t))$ where $Q = L[Y(t_1), Y(t_2), \cdots, Y(t_n)]$. Note that $\Sigma_{j=1}^n \hat{c}_j(t)k(\cdot, t_j) = P_J(k(\cdot, t))$. Let $\tilde{w} = k(\cdot, t)$ and $\tilde{u} = u_{\hat{\lambda}}$. It follows from Lemma 3.1 in the manner of Kimeldorf and Wahba (1970b) that $\hat{w} = \Sigma_{j=1}^n \hat{c}_j(t)k(\cdot, t_j)$, $\hat{u} = \tilde{u} = u_{\hat{\lambda}}$, and

$$\begin{aligned} \hat{y}(t) &= \Sigma_{j=1}^n \hat{c}_j(t)\hat{\lambda}_j \\ &= \Sigma_{j=1}^n \hat{c}_j(t)\langle \tilde{u}, k(\cdot, t_j)\rangle \\ &= \langle \tilde{u}, \Sigma_{j=1}^n \hat{c}_j(t)k(\cdot, t_j)\rangle \\ &= \langle \tilde{u}, \hat{w}\rangle \\ &= \langle \hat{u}, \tilde{w}\rangle \\ &= \langle u_{\hat{\lambda}}, k(\cdot, t)\rangle \\ &= u_{\hat{\lambda}}(t). \end{aligned}$$

REMARK 4.0. If $\{Y(t) : t \in I\}$ is Gaussian, then for fixed $t \in I$, the Theorem 4.1 prediction $\hat{y}(t) = \Sigma_{j=1}^n \hat{c}_j(t)\hat{\lambda}_j$ for $Y(t)$ is the maximum likelihood prediction for $Y(t)$ based on $\mathbf{Y} \in S$. That is, $\hat{y}(t)$ maximizes the joint density of $[Y(t), Y(t_1), \cdots, Y(t_n)]$ subject to $[Y(t_1), Y(t_2), \cdots, Y(t_n)]' \in S$.

REMARK 4.1. If S is a closed subset of R^n, then there exists an n-vector $\hat{\lambda}$ minimizing (4.1) for $\lambda \in S$. If S is also convex, then $\hat{\lambda}$ is the unique n-vector minimizing (4.1) for $\lambda \in S$. If the hypotheses of Theorem 4.1 are satisfied and if there exists a unique n-vector $\hat{\lambda}$ minimizing (4.1) for $\lambda \in S$, then the prediction function $\hat{y}$ of Theorem 4.1 is the unique function y minimizing $\|y\|^2_{H(k)}$ among the functions satisfying (a) and (b) of Theorem 4.1.

In the following example the prediction problem examined by Kimeldorf and Wahba (1970a) is altered by assuming that observations are imprecise rather than precise.

EXAMPLE 4.1. Let $\mathcal{L}$ be of the form (2.3). Let H_a denote $\{f : f \in L^2(-\infty, \infty)$, $\mathfrak{D}^{q-1}f$ is absolutely continuous on compact subintervals of $(-\infty, \infty)$ and $\mathcal{L} f \in L^2(-\infty, \infty)\}$. Let $\{Y(t) : t \in (-\infty, \infty)\}$ be a Gaussian process with mean zero and covariance kernel k where k satisfies $\{f \in H(k)\} = H_a$ and $\mathcal{L}$ maps $H(k)$ isometrically onto $L^2(-\infty, \infty)$. (See Kimeldorf and Wahba (1970a) for details.) Suppose that the values of $Y(t_1), Y(t_2), \cdots, Y(t_n)$ are not known precisely but for some $S \subseteq R$, it is observed that $\mathbf{Y} \in S$. Suppose that S is a closed subset of R^n; hence, there exists $\hat{\lambda} \in S$ minimizing (4.1) for $\lambda \in S$. Since the process is Gaussian, it follows from Lemma 4.1 that $\hat{\lambda}$ is a maximum likelihood estimate for the unknown vector value λ^0 of $\mathbf{Y}$. For $t \in (-\infty, \infty)$, let

$$\Sigma_{j=1}^n \hat{c}_j(t)Y(t_j) = E[Y(t) | Y(t_1), Y(t_2), \cdots, Y(t_n)],$$

and let

$$\hat{y}(t) = \Sigma_{j=1}^{n} \hat{c}_j(t)\hat{\lambda}_j.$$

Then the conditional expectation prediction function $\hat{y}$ is an $\mathfrak{L}$-spline of interpolation to S. That is, $\hat{y}$ minimizes $\int_{-\infty}^{\infty}[(\mathfrak{L}y)(t)]^2\, dt$ among functions satisfying $y \in H_a$ and $[y(t_1), y(t_2), \cdots, y(t_n)]' \in S$.

REMARK 4.1. If, in the previous example, there exists a unique element $\hat{\boldsymbol{\lambda}}$ minimizing (4.1) for $\boldsymbol{\lambda} \in S$, then the conditional expectation prediction function $\hat{y}$ is the unique $\mathfrak{L}$-spline of interpolation to the set S.

REMARK 4.2. Suppose that the mean function m_0 for the process $\{Y(t) : t \in I\}$ is assumed to be known, but it is not assumed that $m_0(t) \equiv 0$. A prediction function can be constructed and characterized for this process; the construction and characterization correspond to the special case $M = \{m_0\}$ of Section 6.

5. Unbiased linear prediction and mean estimation. In this section, $\{Y(t) : t \in I\}$ is assumed to have the model (1.1) where M is a finite-dimensional linear space of functions on I. Let $M_2 = M \cap H(k)$ and let M_1 satisfy $M_1 \cap M_2 = \{0\}$ and $M_1 + M_2 = M$. Let $\{f_i : i = 1, 2, \cdots, q\}$ be a basis for M_1, and let $k_1(s, t) = \Sigma_{i=1}^{n} f_i(s)f_i(t)$. Let $k_0 = k_1 + k$. Then $\{f : f \in H(k_1)\} = M_1$ and $H(k_0)$ is the orthogonal sum of $H(k_1)$ and $H(k)$. For elements $t_1, t_2, \cdots, t_n$ of I, let $J_0 = L[k_0(\cdot, t_j) : j = 1, 2, \cdots, n]$. Let $\mathbf{Y} = [Y(t_1), Y(t_2), \cdots, Y(t_n)]'$ and $\mathbf{X} = [X(t_1), X(t_2), \cdots, X(t_n)]'$. For each $m \in M$, let $\mathbf{m} = [m(t_1), m(t_2), \cdots, m(t_n)]'$. Let $\mathbf{M} = \{\mathbf{m} = [m(t_1), m(t_2), \cdots, m(t_n)]' : m \in M\}$. Note that observing that $\mathbf{Y} \in S$, given the knowledge that m_0 is an unknown element of M, is equivalent to observing that $\mathbf{X} \in S - \mathbf{M}$. The procedure for estimating the n-vector $\boldsymbol{\lambda}^0$ of values taken on by $\mathbf{Y}$ will be not to estimate $\boldsymbol{\lambda}^0$ directly, but to estimate the unknown n-vector $\boldsymbol{\lambda}^0 - \mathbf{m}_0$ of values taken on by $\mathbf{X}$ in the manner of Section 4.

The condition

$$P_M(J_0) = M \tag{5.1}$$

is equivalent to the condition that, for each $t \in I$, there exists an unbiased linear predictor for $Y(t)$. Also, (5.1) implies that each element of $\mathbf{M}$ corresponds to a unique element of M. As before, let $K = [k(t_i, t_j)]_{n \times n}$. Suppose the sets S and $\mathbf{M}$ are such that there exist elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing

$$(\boldsymbol{\lambda} - \mathbf{m})'K^{-1}(\boldsymbol{\lambda} - \mathbf{m}) \tag{5.2}$$

for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$. Then $\hat{\boldsymbol{\lambda}} - \hat{\mathbf{m}}$ estimates the unknown value $\boldsymbol{\lambda}^0 - \mathbf{m}_0$ taken on by $\mathbf{X}$, and $\hat{\boldsymbol{\lambda}}$ estimates $\boldsymbol{\lambda}^0$.

THEOREM 5.1. *Let $\{Y(t) : t \in I\}$ have the model* (1.1) *where M is a finite-dimensional linear space of functions on I. Suppose that it is observed that $\mathbf{Y} \in S$; equivalently, it is observed that $\mathbf{X} \in S - \mathbf{M}$. Assume that* (5.1) *is satisfied. For fixed $t \in I$, let $\hat{Y}(t) = \Sigma_{j=1}^{n} \hat{c}_j(t)Y(t_j)$ be the* MEUL *predictor for $Y(t)$. If there exist* *elements* $\hat{\boldsymbol{\lambda}} \in S$ *and* $\hat{\mathbf{m}} \in \mathbf{M}$ *that minimize* (5.2) *for* $\boldsymbol{\lambda} \in S$ *and* $\mathbf{m} \in \mathbf{M}$, *then the* MEUL *prediction function* $\hat{y}$ *given by* (1.2) *minimizes*

$$\|P_{M^\perp}(y)\|^2_{H(k_0)}$$

among functions satisfying

(a) $y \in H(k_0)$

(b) $[y(t_1), y(t_2), \cdots, y(t_n)]' \in S$.

PROOF. For fixed $t \in I$, if $\tilde{w} = k_0(\cdot, t)$, $H = H(k_0)$, $H_1 = M$, and $H_2 = M^\perp$, then an application of Lemma 3.1 yields an element $\hat{w} = \sum_{j=1}^n \hat{c}_j(t) k_0(\cdot, t_j)$ such that $\sum_{j=1}^n \hat{c}_j(t) Y(t_j)$ is the unique MEUL predictor for $Y(t)$. For any n-vector $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_n]' \in R^n$, it follows from Theorem 3.1 of Peele and Kimeldorf (1977) that the function $\hat{u}_{\boldsymbol{\lambda}}$, where $\hat{u}_{\boldsymbol{\lambda}}(t) = \sum_{j=1}^n \hat{c}_j(t)\lambda_j$ for $t \in I$, is the unique function $u \in H(k_0)$ minimizing $\|P_{M^\perp}(u)\|^2$ subject to the condition $u(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$. It follows from (5.10) of Peele and Kimeldorf (1977) that $P_{M^\perp}(\hat{u}_{\boldsymbol{\lambda}})$ is of the form $\sum_{j=1}^n c_j k(\cdot, t_j)$. Hence if $m_{\boldsymbol{\lambda}} = P_M(\hat{u}_{\boldsymbol{\lambda}})$, then $\|P_{M^\perp}(\hat{u}_{\boldsymbol{\lambda}})\|^2 = (\boldsymbol{\lambda} - \mathbf{m}_{\boldsymbol{\lambda}})' K^{-1} (\boldsymbol{\lambda} - \mathbf{m}_{\boldsymbol{\lambda}})$. It now follows from the minimizing properties of $\hat{u}_{\boldsymbol{\lambda}}$ that

$$\|P_{M^\perp}(\hat{u}_{\boldsymbol{\lambda}})\|^2 = \min_{\mathbf{m} \in \mathbf{M}} (\boldsymbol{\lambda} - \mathbf{m})' K^{-1} (\boldsymbol{\lambda} - \mathbf{m}).$$

Since $\hat{y} = \hat{u}_{\hat{\boldsymbol{\lambda}}}$, the theorem follows.

REMARK 5.0. A special case of Theorem 5.1 is considered by Wahba [4].

REMARK 5.1. If the hypotheses of Theorem 5.1 are satisfied and if the elements $\hat{\boldsymbol{\lambda}}$ and $\hat{\mathbf{m}}$ are the unique elements minimizing (5.2), then the MEUL prediction function $\hat{y}$ of Theorem 5.1 is the unique function y minimizing $\|P_{M^\perp}(y)\|^2$ subject to constraints (a) and (b).

REMARK 5.2. If S is a closed and bounded subset of R^n, then $S - \mathbf{M}$ is closed and, consequently, there exist elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$. If S is a closed subset of R^n and if S and $\mathbf{M}$ are orthogonal with respect to the Hilbert space R^n, then $S - \mathbf{M}$ is closed and there exist elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$.

REMARK 5.3. If S is a closed, convex, bounded subset of R^n and if $\mathbf{M} \cap \{s = s_1 - s_2 : s_1 \in S, s_2 \in S\} = \{0\}$, then there exist unique elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$. If S is a closed, convex subset of R^n such that S and $\mathbf{M}$ are orthogonal subsets of R^n, then there exist unique elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$.

EXAMPLE 5.1. Let $\mathfrak{L}$ be of the form (2.3). Let $H_b = \{f : \mathfrak{D}^{q-1} f$ is absolutely continuous and $\mathfrak{L}f \in L^2[0, 1]\}$, and let $\{f_i : i = 1, 2, \cdots, q\}$ be a basis for the null space of $\mathfrak{L}$. Let M be the null space of $\mathfrak{L}$. Let k be such that $\{f \in H(k)\} = \{f \in H_b : (\mathfrak{D}^i f)(0) = 0$ for $i = 0, 1, \cdots, q - 1\}$, and $\mathfrak{L}$ maps $H(k)$ isometrically onto $L^2[0, 1]$. (See Kimeldorf and Wahba (1970b) for details.) Suppose that

$\{Y(t) : t \in [0, 1]\}$ has the model (1.1) with the above k and M, and let $T = \{t_1, t_2, \cdots, t_n\} \subseteq [0, 1]$ satisfy rank $[f_i(t_j)]_{q\times n} = q$. (This matrix condition is equivalent to condition (5.1).) Suppose that for some $S \subseteq R$, it is observed that $\mathbf{Y} \in S$ where $\mathbf{Y} = [Y(t_1), Y(t_2), \cdots, Y(t_n)]'$. If there exist elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ that minimize (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$, then the MEUL prediction function $\hat{y}$ of Theorem 5.1 is an $\mathfrak{L}$-spline of interpolation to S. That is, $\hat{y}$ minimizes $\int_0^1[(\mathfrak{L}y)(t)]^2\,dt$ among functions y satisfying $y \in H_b$ and $[y(t_1), y(t_2), \cdots, y(t_n)]' \in S$.

REMARK 5.4. If, in Example 5.1, the elements $\hat{\boldsymbol{\lambda}}$ and $\hat{\mathbf{m}}$ are the unique elements of S and $\mathbf{M}$ that minimize (5.2), then the MEUL prediction function $\hat{y}$ is the unique $\mathfrak{L}$-spline of interpolation to S.

THEOREM 5.2. *Suppose that the hypotheses of Theorem* 5.1 *are satisfied. For fixed* $t \in I$, *let* $\hat{Z}(t) = \sum_{j=1}^n \hat{d}_j(t)Y(t_j)$ *be the* MVUL *estimator for* $m_0(t)$. *Then the* MVUL *mean-estimation function* $\hat{z}$, *where* $\hat{z}(t) = \sum_{j=1}^n \hat{d}_j(t)\hat{\lambda}_j$ *for* $t \in I$, *is* $P_M(\hat{y})$ *where* $\hat{y}$ *is the* MEUL *prediction function of Theorem* 5.1.

PROOF. The proof is an immediate consequence of Theorem 3.2 of Peele and Kimeldorf (1977) and Theorem 5.1.

REMARK 5.5. Let $\boldsymbol{\lambda}^0 \in R^n$. Since $\boldsymbol{\lambda}^0 - \mathbf{M}$ is closed and convex, there exists a unique element $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\mathbf{m} \in \mathbf{M}$. Suppose that (5.1) is satisfied. It can be seen that the unique element $m = \hat{m} \in M$ satisfying $[m(t_1), m(t_2), \cdots, m(t_n)]' = \hat{\mathbf{m}}$ is the MVUL mean-estimation function $\hat{z}$ of Theorem 3.2 of Peele and Kimeldorf (1977).

6. Biased prediction functions.

In previous prediction problems with an unknown mean function, M was a linear space and unbiased linear prediction was possible. Suppose now that the process has the model (1.1) with M assumed only to be a set of functions on I such that the linear span of M is finite-dimensional. Suppose that unbiased linear prediction is not possible. As before let $t_1, t_2, \cdots, t_n$ be distinct elements of I, and let $\mathbf{M} = \{[m(t_1), m(t_2), \cdots, m(t_n)]' : m \in M\}$.

Suppose it is observed that $\mathbf{Y} \in S$, or, equivalently, it is observed that $\mathbf{X} \in S - \mathbf{M}$. If there exist elements $\hat{\boldsymbol{\lambda}} \in S$ and $\hat{\mathbf{m}} \in \mathbf{M}$ minimizing (5.2) for $\boldsymbol{\lambda} \in S$ and $\mathbf{m} \in \mathbf{M}$, then any element $\hat{m} \in M$ satisfying $[\hat{m}(t_1), \hat{m}(t_2), \cdots, \hat{m}(t_n)]' = \hat{\mathbf{m}}$ is as good an estimate for m_0 as any. Let $\hat{W} = \sum_{j=1}^n \hat{c}_j(t)X(t_j)$ minimize $E(W - X(t))^2$ for $W \in L[X(t_1), X(t_2), \cdots, X(t_n)]$. Given that $\mathbf{Y}$ is observed to be in S, the random variable $\hat{Y}(t) = \hat{m}(t) + \sum_{j=1}^n \hat{c}_j(t)X(t_j)$ is a predictor for $Y(t)$, and the function $\hat{y}$ where

$$\hat{y}(t) = \hat{m}(t) + \sum_{j=1}^n \hat{c}_j(t)\big(\hat{\lambda}_j - \hat{m}(t_j)\big) \tag{6.1}$$

is a prediction function.

THEOREM 6.1. *Let* $\{Y(t) : t \in I\}$ *have the model* (1.1) *where* M *is a set of functions on* I. *For fixed* $t \in I$, *let* $\sum_{j=1}^n \hat{c}_j(t)X(t_j)$ *be a random variable* W *minimizing*

$E(W - X(t))^2$ for $W \in L[X(t_1), X(t_2), \cdots, X(t_n)]$. Suppose that for some $S \subseteq R^n$, $\mathbf{Y} \in S$ is observed. If there exists an element $\hat{m} \in M$ and an element $\hat{\boldsymbol{\lambda}} \in S$ such that $\hat{\mathbf{m}}$ and $\hat{\boldsymbol{\lambda}}$ minimize (5.2) for $\mathbf{m} \in \mathbf{M}$ and $\boldsymbol{\lambda} \in S$, then the prediction function $\hat{y}$ given by (6.1) satisfies

$$\|\hat{y} - \hat{m}\|^2_{H(k)} \leqslant \|y - m\|^2_{H(k)}$$

for all functions $y \in M + H(k)$ and $m \in M$ satisfying

(a) $y - m \in H(k)$

and

(b) $[y(t_1), y(t_2), \cdots, y(t_n)]' \in S$.

PROOF. Let $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \cdots, \lambda_n]' \in R^n$. For $t \in I$, let $\hat{u}_{\boldsymbol{\lambda}}(t) = \Sigma_{j=1}^n \hat{c}_j(t)\lambda_j$. Hence $\hat{u}_{\boldsymbol{\lambda}}$ is the unique element of $H(k)$ of minimal norm among functions interpolating the points $(t_1, \lambda_1), (t_2, \lambda_2), \cdots, (t_n, \lambda_n)$. It follows that $\hat{y} - \hat{m} \in H(k)$ and $\hat{y} \in M + H(k)$. Condition (b) follows from (6.1). The theorem follows from the observation that for $\boldsymbol{\lambda} \in R^n$, $\hat{u}_{\boldsymbol{\lambda}} = \Sigma_{j=1}^n c_j k(\cdot, t_j)$ where $\mathbf{c} = [c_1, c_2, \cdots, c_n]'$ satisfies $K\mathbf{c} = \boldsymbol{\lambda}$ and hence $\|\hat{u}_{\boldsymbol{\lambda}}\|^2 = \mathbf{c}'K\mathbf{c} = (K^{-1}\boldsymbol{\lambda})K(K^{-1}\boldsymbol{\lambda}) = \boldsymbol{\lambda}'K^{-1}\boldsymbol{\lambda}$.

REMARK 6.1. Let $S = \{\boldsymbol{\lambda}^0\}$ for some $\boldsymbol{\lambda}^0 \in R^n$, and let M be a finite-dimensional linear space of functions on I. Since $\boldsymbol{\lambda}^0 - \mathbf{M}$ is a closed, convex subset of R^n, there exists a unique element $\hat{\mathbf{m}} \in \mathbf{M}$ which minimizes (5.2) for $\mathbf{m} \in \mathbf{M}$. Let k_0 be defined as in Section 5. If condition (5.1) is satisfied, then for each $t \in I$ there exists an unbiased linear predictor for $Y(t)$ and the prediction function $\hat{y}$ of Theorem 6.1 is the same as the MEUL prediction function of Theorem 3.1 of Peele and Kimeldorf (1977). If condition (5.1) is not satisfied, then unbiased linear prediction for $Y(t)$ for all $t \in I$ is not possible. Also the prediction function $\hat{y}$ of Theorem 6.1 will not be unique if condition (5.1) is not satisfied since there will exist a nonzero element $m \in M$ such that $\mathbf{m}$ is the zero n-vector.

EXAMPLE 6.1. Suppose that the prediction problem of Kimeldorf and Wahba (1970b) is altered by assuming that rank $[f_i(t_j)]_{q \times n} < q$; so, unbiased linear prediction is not possible. It follows from Remark 6.1 that the prediction function of Theorem 6.1 is a nonunique $\mathfrak{L}$-spline of interpolation to the points $\{(t_j, \lambda_j) : j = 1, 2, \cdots, n\}$.

7. **Infinitely many imprecise observations.** Let $\{Y(t) : t \in I\}$ have the model (1.1) with mean zero. A new look at the prediction problem examined in Section 4 provides a method for predicting with infinitely many imprecise observations. In Section 4 it is assumed that $T = \{t_1, t_2, \cdots, t_n\}$, and $\mathbf{Y} = [Y(t_1), Y(t_2), \cdots, Y(t_n)]'$ is observed to be in some known set S of n-vectors. The unknown value $\boldsymbol{\lambda}^0$ of $\mathbf{Y}$ is estimated by a vector $\hat{\boldsymbol{\lambda}}$ that minimizes $\boldsymbol{\lambda}'K^{-1}\boldsymbol{\lambda}$ for $\boldsymbol{\lambda} \in S$. Let $u_{\boldsymbol{\lambda}} = \Sigma_{j=1}^n c_j k(\cdot, t_j)$ where $\mathbf{c} = K^{-1}\boldsymbol{\lambda}$ and $[c_1, c_2, \cdots, c_n]' = \mathbf{c}$. Then $u_{\boldsymbol{\lambda}} \in H(k)$, $u_{\boldsymbol{\lambda}}(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$, and $\|u_{\boldsymbol{\lambda}}\|^2 = \boldsymbol{\lambda}'K^{-1}\boldsymbol{\lambda}$. Let k_T be the restriction of the function k on $I \times I$ to the set $T \times T$. Then $k_T(\cdot, t_j)$ is a function on T, and the function $u_{\boldsymbol{\lambda}, T}$ on T defined by $u_{\boldsymbol{\lambda}, T} = \Sigma_{j=1}^n c_j k_T(\cdot, t_j)$ satisfies

$u_{\lambda, T} \in H(k_T)$, $u_{\lambda, T}(t_j) = \lambda_j$ for $j = 1, 2, \cdots, n$, and $\|u_{\lambda, T}\|^2_{H(k_T)} = \lambda' K^{-1}\lambda$. Hence, observing that $\mathbf{Y} \in S$ is equivalent to observing that the sample function for $\{Y(t) : t \in T\}$ is in the set s where $s = \{f \in H(k_T) : [f(t_1), f(t_2), \cdots, f(t_n)]' \in S\}$, and choosing $\hat{\lambda}$ to minimize $\lambda' K^{-1}\lambda$ for $\lambda \in S$ is equivalent to choosing $\hat{f}$ to minimize $\|f\|^2_{H(k_T)}$ for $f \in s$.

For the case when T contains infinitely many elements, the natural extension of the previous procedure is to assume that the sample function for $\{Y(t) : t \in T\}$ is known to belong to a known subset s of $H(k_T)$ such that there exists $\hat{f} \in s$ which minimizes $\|f\|^2_{H(k_T)}$ for $f \in s$. Consider the case when T is a sequence $\{t_j\}_{j=1}^{\infty}$. Let $\{\tilde{Y}_j\}_{j=1}^{\infty}$ be the Gram-Schmidt orthonormalization of $\{Y(t_j)\}_{j=1}^{\infty}$; that is, for a positive integer n, $\tilde{Y}_n = Y_n/\|Y_n\|$ where

$$Y_n = Y(t_n) - \Sigma_{j=1}^{n-1}\langle Y(t_n), \tilde{Y}_j\rangle \tilde{Y}_j.$$

Hence, $\{\tilde{Y}_j\}_{j=1}^{\infty}$ is an orthonormal basis for $L^2[Y(t) : t \in T]$. Let $f \in H(k_T)$, and let F be the isometric image in $L^2[Y(t) : t \in T]$ of f under the isometry taking $k_T(\cdot, t_j)$ to $Y(t_j)$. Then f is the sample function for $\{Y(t) : t \in T\}$ if and only if for each positive integer n, $Y(t_n) = f(t_n) = \langle f, k_T(\cdot, t_n)\rangle = \langle F, Y(t_n)\rangle$. Equivalently, f is the sample function for $\{Y(t) : t \in T\}$ if and only if for each positive integer n, $\tilde{Y}_n = \langle F, \tilde{Y}_n\rangle$. Since $F = \Sigma_{j=1}^{\infty}\langle F, \tilde{Y}_j\rangle \tilde{Y}_j$ and $\|f\|^2_{H(k_T)} = \|F\|^2$, it follows that $\|f\|^2_{H(k_T)}$ is the sum $\Sigma_{j=1}^{\infty}\langle F, \tilde{Y}_j\rangle^2$ of the squares of the values of the sequence $\{\tilde{Y}_j\}_{j=1}^{\infty}$ of orthonormal random variables. Thus, estimating the sample function for $\{Y(t) : t \in T\}$ by a function $\hat{f}$ which minimizes $\|f\|^2_{H(k_T)}$ for $f \in s$ is justified statistically. For $W \in L^2[Y(t) : t \in T]$, if w is the isometric image in $H(k_T)$ of W under the usual isometry, then it follows from the previous sentence and the paragraph immediately preceding Theorem 5.1 of Peele and Kimeldorf (1977) that $\langle \hat{f}, w\rangle_{H(k_T)}$ estimates W. For fixed $t \in I$, let $\hat{Y}(t) \in L^2[Y(t) : t \in T]$ be a predictor for $Y(t)$, and let $\hat{w}_t$ be the isometric image in $H(k_T)$ of $\hat{Y}(t)$. Then the function $\hat{y}$ defined by

$$\hat{y}(t) = \langle \hat{f}, \hat{w}_t\rangle_{H(k_T)} \quad \text{for} \quad t \in I \tag{7.1}$$

is a prediction function.

THEOREM 7.1. *Let $\{Y(t) : t \in I\}$ have the model* (1.1) *with $m_0 \equiv 0$. Let $T = \{t_j\}_{j=1}^{\infty}$. For fixed $t \in I$, let $\hat{Y}(t)$ be a random variable W minimizing $E(W - Y(t))^2$ for $W \in L^2[Y(t) : t \in T]$. Suppose that the sample function for $\{Y(t) : t \in T\}$ is known to belong to a known subset s of $H(k_T)$, and suppose that $\hat{f}$ minimizes $\|f\|^2_{H(k_T)}$ for $f \in s$. The prediction function $\hat{y}$ given by* (7.1) *minimizes*

$$\|y\|^2_{H(k)}$$

among functions satisfying

(a) $y \in H(k)$

and

(b) $y_T \in s$

where y_T is the restriction to T of y.

PROOF. Let $J = L^2[k(\cdot, t) : t \in T]$. Then $L^2[k_T(\cdot, t) : t \in T]$ is isometric to J under the linear mapping $\mathfrak{F}$ that takes $k_T(\cdot, t)$ to $k(\cdot, t)$ for each $t \in T$, and the quantity $\langle \hat{f}, \hat{w}_t \rangle_{H(k_T)}$ of (7.1) is equal to $\langle \mathfrak{F}(\hat{f}), \mathfrak{F}(\hat{w}_t) \rangle_{H(k)}$. For $f \in L^2[k_T(\cdot, t) : t \in T]$, $\mathfrak{F}(f)$ is the unique function $u \in H(k)$ that minimizes $\|u\|^2$ subject to the constraints $u(t) = f(t)$ for $t \in T$. Since $\|\mathfrak{F}(f)\|_{H(k)} = \|f\|_{H(k_T)}$, $\mathfrak{F}(\hat{f})$ satisfies the conditions of the theorem. It remains only to show that the prediction function $\hat{y}$ is the function $\mathfrak{F}(\hat{f})$. For fixed $t \in I$, let $\tilde{w} = k(\cdot, t)$. Now apply Lemma 3.1 with $\tilde{u} = \mathfrak{F}(\hat{f})$, $H_1 = \{0\}$, and $H = H_2 = H(k)$. The function $\hat{u}$ of Lemma 3.1 is $\tilde{u}$ and the isometric image in $H(k)$ of $\hat{Y}(t)$ is $\hat{w}$. Since $\hat{y}(t) = \langle \hat{f}, \hat{w}_t \rangle_{H(k_T)}$ and $\hat{w} = \mathfrak{F}(\hat{w}_t)$, it follows that

$$\begin{aligned} \hat{y}(t) &= \langle \mathfrak{F}(\hat{f}), \hat{w} \rangle_{H(k)} \\ &= \langle \tilde{u}, \hat{w} \rangle_{H(k)} \\ &= \langle \hat{u}, \tilde{w} \rangle_{H(k)} \\ &= \langle \mathfrak{F}(\hat{f}), k(\cdot, t) \rangle_{H(k)} \\ &= [\mathfrak{F}(\hat{f})](t). \end{aligned}$$

Hence, $\hat{y} = \mathfrak{F}(\hat{f})$ and the theorem follows.

EXAMPLE 7.1. Let $\mathfrak{L}$ be the derivative operator. If Example 5.1 of Kimeldorf and Peele (1977) is altered to fit the hypotheses of Theorem 7.1, then the prediction function $\hat{y}$ of Theorem 7.1 will minimize $\int_0^1 [(\mathfrak{L}y)(t)]^2 \, dt$ among functions y satisfying $y \in H_b$ and constraint (b) of Theorem 7.1.

REFERENCES

[1] KIMELDORF, G. S. and WAHBA, G. (1970a). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* **41** 495–502.

[2] KIMELDORF, G. S. and WAHBA, G. (1970b). Spline functions and stochastic processes. *Sankhyā* **32** 173–180.

[3] PEELE, L. and KIMELDORF, G. (1977). Prediction functions and mean-estimation functions for a time series. *Ann. Statist.* **5** 709–721.

[4] WAHBA, G. (1969). Estimating derivatives from outer space. Mathematics Research Center Report No. 989, Univ. Wisconsin, Madison.

DEPARTMENT OF MATHEMATICAL AND COMPUTING SCIENCES
OLD DOMINION UNIVERSITY
NORFOLK, VIRGINIA 23508

PROGRAMS IN MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
BOX 688
RICHARDSON, TEXAS 75080