

SMOOTHING LOCALLY SMOOTH PROCESSES  
BY BAYESIAN NONPARAMETRIC METHODS

by

S. Weerahandi &  
J.V. Zidek

Technical Report No. 26

December, 1985

ABSTRACT

This report is concerned with the analysis of stochastic processes of the form  $R(x) = S(x) + N(x)$  where  $S$  is a "smooth",  $N$  is noise, and all quantities including  $x$  are real-valued. Our methods, which can easily be extended to a much more general class of stochastic processes, derive from the assumption that the observed  $R$ -values and unobserved values of  $R$  (interpolants or predictands, the assumed inferential objectives of the analysis), are linearly related through Taylor series expansions of observed about unobserved values. The expansion errors and all other a priori unspecified quantities have a joint multivariate normal distribution which expresses the prior uncertainty about their values. For technical expediency, both maximum likelihood and cross-validation are used to estimate the a priori unspecified hyperparameters. The results include interpolators, predictors and derivative estimates with credibility interval estimates automatically generated in each case. An analysis of an acid rain, wet-deposition time series is included to indicate the efficacy of the proposed method.

KEYWORDS. Non-parametric regression; acid rain; Bayesian regression; time series; spatial interpolation; Kriging; forecasting; prediction.

INTRODUCTION AND SUMMARY. This report is concerned with interpolation and prediction for stochastic processes. Although our analysis focuses on a special case, the methods can easily be extended to a very general class of multiple, multivariate space-time-regression models. The structure of the response vectors for the most general class, as a function of a vector argument of space, time and regressor variable co-ordinates, is given by  $R(x) = M(x) + S(x) + N(x)$ . Here  $N$  is an independent noise process,  $S$  is a "smooth", a locally Taylor expandable function, and  $M$  is a model, possibly with a priori unspecified parameters. In this paper, to avoid obscuring the basic ideas, a particular case is studied where  $M \equiv 0$  and all vectors are real, that is, one dimensional. However an extension of our results to the general case is straightforward.

We assume  $R$  has been observed at a sequence of  $x$ -values,  $t_1, \dots, t_n$ , and that the objective of inference is the estimation of some unobserved  $R$  value. In other words, the goal is either interpolation or extrapolation of  $\alpha = R(t)$  at a point,  $x = t_{n+1}$  at which  $R$  has not yet been observed. (Credibility) interval estimations of  $\alpha$  are deemed to be necessary.

The key components of our analysis are firstly, the expansions to finite order,  $S(t_i) = \alpha + \beta(t_i - t_{n+1}) + \dots + \epsilon_i$  where  $\beta = dS(t_{n+1})/dx$ , and  $\epsilon_i$  denotes the remainder in the Taylor expansion,  $i=1, \dots, n$ . And secondly, we assume that all a priori unspecified quantities like  $\alpha$ ,  $\beta$ , and the  $R(t_i)$  have a joint distribution which we will take to be multivariate normal; all our results then follow quite simply. However, any joint distribution which seemed appropriate could, in principle, have been used instead of the normal.

In Section 2 the theory underlying our approach is outlined. In Section 3 we illustrate its use in the analysis of an acid rain wet deposition time series. Our method is then compared and contrasted in Section 4 with other approaches which have been proposed for the analysis of nonparametric series. Section 5 includes a brief discussion and our conclusions.

The idea for this work derived from conversations in 1978 with Professor N. Glick in which he described a running percentile approach to smoothing data and generating reliability bands. While his analysis was data-analytical in character and concerned specifically with the assessment of obesity in children, it did suggest the method described here as a way of explicating his methods. We are indebted to Dr. Glick for his stimulating comments and active encouragement in our work.

ACKNOWLEDGEMENTS. We are very grateful to Wilson (Hon) Ma for his considerable assistance with the numerical work described in Section 3. The final presentation benefitted greatly from the comments of Dr.'s Harry Joe and A. John Petkau. This work was partially supported by the Environmental Protection Agency through a co-operative research agreement with SIAM's Institute for Mathematics and Society (SIMS). It was also supported by the National Science Foundation of the National Science Foundation of the United States (Grant #MCS83-01807) and the Natural Science and Engineering Research Council of Canada.

## 2. LOCALLY SMOOTH PROCESSES

### 2.1 PREAMBLE

The report is concerned with stochastic processes which have smooth underlying "carriers". An example would be  $R = R(t) = R(t, \omega)$ , the measured weight of an individual  $\omega$  at time  $t$ .  $R$  would vary between successive weighings because of rounding errors, imprecision of the weighing scale and other factors, some of which would be undetermined. But there is a purely conceptual, underlying carrier function  $C = C(t) = C(t, \omega)$  which represents  $\omega$ 's intrinsic weight and embodies intermediate to long term trends. In a frequency-theory setting,  $C$  would represent the mean or expected weight after all extraneous factors have been "averaged out" over repeated weighings. The impact of such extraneous variation is embodied in  $R - C = N = N(t) = N(t, \omega)$ , the "noise" term, which here, and in general would be of little intrinsic interest.

In the weighing example,  $N$  would be almost negligible as a percentage of  $R$ . In contrast, for the application considered in Section 3,  $N$  would be appreciable. There  $R$ , the measured acidity or pH level of wet precipitation, varies considerably even during a given precipitation event.  $R$  is of interest as an indicator of the state of the environment,  $C$ , at the time of the event. Thus, in this application, the role of  $C$  would be analogous to the intensity function in a point process, while  $R$  would be measured only at the times of the events in that process.

The two examples described above differ in other ways as well. In experiments where weight was an important consideration, there would

usually be several randomly selected subjects,  $\omega_1, \dots, \omega_m$ , and the  $R(t, \omega_i)$ ,  $i=1, \dots, m$ ,  $-\infty < t < \infty$  would be regarded as conditionally independent given the  $\{C(t, \omega)\}$ . Usually  $C(t, \omega_i) = C(t, \omega_j)$ ,  $i \neq j$ , would be assumed when all other explanatory variables had been accounted for. And  $R$  might be recorded at just a single time point,  $t = t_i$  for subject  $i$ .

On the other hand, in the acid rain example, there is just a single path,  $R(t, \omega)$ , and observations will be recorded at a sequence of time points. In a full treatment of this problem  $R(t, \omega)$ , would be a multivariate series  $R(x, \omega)$ , where  $R$  is a vector of measurements like the average pH level during the precipitation event, the volume of precipitation, the average concentration of sulphate, and so on. Here  $x = (x_1, x_2, \dots, x_k)$  might be the space-time co-ordinates (including elevation) at which the measurements are made. Or  $x$  might be a vector of independent regressors like distance to the nearest major source of airborne contaminants.

Although our conceptual model embraces a wide variety of situations, the methodology we propose in this report is very general and may be adapted to permit inferences to be made in all of the situations we have envisaged. This report focuses on the case of a single, real-valued series,  $R = R(t)$  where  $t$  is real, an element of an open subset of the real line;  $t$  represents "time" in Section 3.

Interpolation and prediction with attendant measures of reliability are the objectives of our analysis. Measurements  $r_i = R(t_i)$ ,  $i = 1, \dots, n$  are assumed to be available. Inference is to be about  $\beta_0 = C(t_{n+1})$ .

In general,  $C = M+S$  where  $M$  is a model for the carrier and  $S$  is a "smooth". The model would incorporate any known structure for  $C$ . It might have jumps, for example, to incorporate discontinuous changes at possibly unknown  $t$ -points.  $M$  might also include a trigonometric series to account for seasonality. Any model with only a finite number of unspecified parameters can readily be incorporated into a conceptually straightforward extension of the results reported here. However, for simplicity it will be assumed here that  $M \equiv 0$ , and that any known structure like linear trend, for example, has been removed.

## 2.2. STRUCTURAL MODELS FOR S

$S$  is not assumed to have a parametric form. Instead it is assumed to be locally regular, that is expandable in a Taylor series about  $t = t_{n+1}$ . A local parametrization of  $S$  is thereby achieved where the parameters are  $\beta_i = D^i S(t_{n+1})$ ,  $D$  denoting the differentiation operator. A structural model for the data is thereby achieved. A priori it is

$$R = XB + \epsilon \quad (2.1).$$

Here  $R = (R_1, \dots, R_n)^T$ ,  $R_i = R(t_i)$ ,  $\beta = (\beta_0, \dots, \beta_p)^T$ ,  $X$ , an  $n \times (p+1)$  matrix, is given by  $X = (1, X_1, \dots, X_p)$ ,  $1$  is an  $n$ -vector of 1's,  $X_j^T = ([t_i - t_{n+1}]^j / j!, \dots, [t_i - t_{n+1}]^j / j!)$ ,  $j = 1, \dots, p$ ,  $\epsilon = \eta + N$ ,  $N^T = (N_1, \dots, N_n)$ ,  $N_i = N(t_i)$ ,  $\eta^T = (\eta_1, \dots, \eta_n)$ , and  $\eta_i$  is the remainder of the Taylor expansion of  $S(t_i)$ , that is  $\eta_i = [t_i - t_{n+1}]^{p+1} D^{p+1} S(\theta_i) / (p+1)!$  where  $\theta_i$  is a point in the interval joining  $t_i$  and  $t_{n+1}$ . In fact this expansion would be valid if  $S$  were deemed to have only  $p$  derivatives but  $\eta_i$  would then be represented otherwise. For simplicity of exposition this form will be assumed.

The degree of local regularity is reflected in the size of  $p$ ; its choice is subjective. If  $S$  is deemed to have  $p+1$  derivatives, it is desirable to include the maximal number of terms,  $p$  in the expansion, (2.1), to reduce the size of the error of approximation,  $\eta_1$ .

### 2.3 STOCHASTIC MODELS FOR S

Although  $R$  is observed  $\beta$ , (see equation (2.1)) and  $c$  are not. The second key element in our approach is the assumption that all uncertainty about  $R$  (before sampling),  $\beta$  and  $c$  is measurable and representable in terms of a (multivariate) probability distribution. Here it is assumed that the co-ordinates of  $N$  are independent, identically distributed random variables which are independent of  $\alpha$  and  $\beta$ . As well, it is assumed that  $(S^T, \beta^T)^T$  has a joint distribution where  $S = (S_1, \dots, S_n)^T$ ,  $S_i = S(t_i)$ ,  $i = 1, \dots, n$ . Let  $\text{Var}(N_1) = \sigma^2$ .

Finding a coherent, probabilistic expression of uncertainty about the unspecified elements of equation (2.1) is not straightforward. It seems reasonable to regard the Taylor series remainders,  $\eta_1$ , as independent with mean zero. And clearly the variances of the  $\eta_1$  should be an increasing function of the  $\{|t_i - t_{n+1}|\}$ , say  $h(|t_i - t_{n+1}|)$ . If  $D^{p+1}S(0)$  is regarded as bounded, the mathematical form of  $\eta_1$  suggests  $h(x) = k^2 x^{2p+2}$  as a first approximation. However this can only be regarded as an approximation since it is unbounded as  $x = |t_i - t_{n+1}| \rightarrow \infty$  whereas the uncertainty about  $\eta_1$ , which is expressed by  $\text{Var}(\eta_1)$ , is necessarily bounded. After all the pH values of rainfall are in the neighbourhood of 5.6.

It is also clear that  $\eta_1$  will depend upon  $\beta$  to an extent which diminishes as  $x$  increases. Consider the case where  $p = 0$ , i.e.  $S(t_1) = S(t_{n+1}) + \eta_1$ . Since the prior distribution fixes the scale of  $S(t_{n+1}) = \beta_0$ , an extremely small value of  $S(t_{n+1})$ , for example, would entail  $\eta_1 > 0$  if  $x$  were small.

After recentering  $\beta$ , if necessary, and the removal of known trends in  $S$ , we would have  $E(\beta) = 0$ ; this will now be assumed. The covariance of  $\beta$ , say  $\Sigma$ , is more complex. Assume  $\text{Var}(S(t)) = \delta^2$  for all  $t$ . Consider  $\text{Cov}(S(t_{n+1}), DS(t_{n+1}))$ , for example. If  $\beta_0$  were, conditionally, an extreme value of  $S$ , this would indicate that  $S$  was at a maximum or minimum and hence  $\beta_1 = 0$ ; if  $\beta_0$  were not extreme nothing could be inferred about  $\beta_1$  so its independence from  $\beta_0$  would be immediate. Hence  $\beta_0$  and  $\beta_1$  are uncorrelated. But similar reasoning about  $\beta_0$  and  $\beta_2$  suggests a possibly strong negative correlation.

Let  $c^*(t,u) = E S(t)S(u)$  denote  $S$ 's auto-covariance function. Assume  $c^*(t,u) = c(t-u)$ . Then as is well-known  $c(-u) = c(u)$ . Assume  $c$  is  $2p$ -fold differentiable. Observe that  $Dc(-t) = \lim[(c(-t)-c(-t-h))/h : h \rightarrow 0+] = -\lim[(c(t+h) - c(t))/h : h \rightarrow 0+] = -Dc(t)$ . It follows that  $Dc(0) = 0$  and by induction that all of odd  $c$ 's derivatives of odd order are zero at  $t=0$  since  $D^{2M+1}c(0) = \lim[(D^{2M-1}c(h)+D^{2M-1}c(-h) - 2D^{2M-1}c(0))/h^2 : h \rightarrow 0+] = 0$  by the induction hypothesis.

Note that  $(-1)^j D^{2j} c(0) \geq 0$  for all  $j$ . To see this first observe that

$$E D^i S(t) D^j S(u) = (-1)^j c^{(i+j)}(t-u), i, j = 0, \dots, p,$$

where, in general  $c^{(r)}(x)$  denotes the  $r$ -th derivative of  $c$  evaluated at  $x$  (Priestly 1981, Section 3.6.4 presents this result when  $i=0, j=1$ ). Taking

the limit in this last result as  $u \rightarrow t$  with  $i=j$  gives the result.

From the analysis presented above, the following representations are obtained, where  $\Sigma = \text{Cov}(\beta) = (\Sigma_{ij})$ :

$$\begin{aligned}\Sigma_{ij} &= (-1)^j c^{(i+j)}(0) \\ \Gamma^j &= (-1)^j c^{(j)}(t-t_{n+1})\end{aligned}\quad (2.2)$$

for  $i, j = 0, \dots, p$  where  $\Gamma^j$  is the  $j$ -th co-ordinate of  $\Gamma = \text{Cov}(S(t), \beta^T)$ , a  $(p+1)$  dimensional row vector for any  $t$ . Observe that for small values of

$$|t - t_{n+1}|, \Gamma^j = (-1)^j \sum_r (t-t_{n+1})^r c^{(r+j)}(0)/r! = \sum_r (t-t_{n+1})^r \cdot \Sigma_{rj} / r!$$

Thus

$$\Gamma = \sum_r (t-t_{n+1})^r \Sigma_r / r! \quad (2.3)$$

where  $\Sigma_r$  denotes the  $r$ th row of  $\Sigma$ .

Now suppose  $(S(t), \beta^T)^T$  has a distribution with  $S(t) = a\beta + \zeta$  where  $a$  is a vector of constants,  $\text{Cov}(\zeta, \beta) = 0$ , and  $E(\zeta) = 0$ . It follows immediately that  $\Gamma = \text{Cov}(S, \beta^T) = a\Sigma$  so  $a = \Gamma\Sigma^{-1}$ . Also  $\text{Var}(\zeta) = c(0) - \Gamma\Sigma^{-1}\Gamma^T$ . In summary we will assume, in conflict with equation (2.1), that

$$R = X^* \beta + \epsilon^* \quad (2.4)$$

where  $X^*$ 's  $i$ -th row is  $\Gamma(t_i)\Sigma^{-1}$ ,  $\Gamma(t_i) = \text{Cov}(S(t_i), \beta^T)$  and  $\epsilon^* = \eta^* + N$ ;  $N$  is the uncorrelated noise process in (2.1),  $\eta_i^* = (c(0) - \Gamma(t_i)\Sigma^{-1}\Gamma^T(t_i))^{1/2} z_i$ , and  $z_i = z_i(t, t_{n+1})$  has mean 0, variance 1 and is uncorrelated with the other components of the righthand side of equation (2.4).

## 2.4 STOCHASTIC VERSUS STRUCTURAL LOCAL MODELS FOR S.

Both of equations (2.1) and (2.4) are exact under appropriate circumstances. At least locally they are approximately the same. To see this observe that equation (2.3) entails  $\Gamma(t_i) \sum_r^{-1} = \sum_r e_r (t_i - t_{n+1})^r / r!$  where  $e_r$  denotes the vector all of whose elements are 0 except the r-th which is 1. Thus  $\Gamma(t_i) \sum_r^{-1} \beta = \sum_r \beta_r (t_i - t_{n+1})^r / r!$  so that the i-th row of  $X^*$  is approximately equal to that of  $X$  when  $t_i$  is near  $t_{n+1}$ . At the same time  $\Gamma(t_i) \sum_r^{-1} \Gamma^T(t_i) = \sum_r (t_i - t_{n+1})^r \Gamma_r / r! = \sum_k x^k c^{(k)}(0) / k! \sum_r \binom{k}{r} (-1)^{k-r} - e = c(0) - e$  where  $x = t_i - t_{n+1}$  and  $e$  denotes the remainder and involves terms of order at least  $x^{p+1}$ . To obtain the correct order would require expanding  $\Gamma(t_i)$  to an order higher than  $p$ . This would require derivatives of  $c$  of order higher than the assumed  $2p$ . And the definition of  $\sum_r$  would have to be formally extended to values of  $r$  exceeding  $p$  using the first of equations (2.2). If these extensions are assumed and only terms of order  $x^{2p+2}$  or lower order were retained, we would have the remainder,  $e$ , given by

$$e = - \sum_{k=p+1}^{2p+2} (x^k / k!) \sum_{r=0}^k \binom{k}{r} \sum_r^{-1} \sum_r^{-1} \sum_{k-r}^{-1}$$

$$= -2 \sum_{k=p+1}^{2p+2} (x^k / k!) \sum_{r=0}^{k-p-1} \binom{k}{r} \left[ \sum_r^{-1} \sum_{k-r}^{-1} - c^{(k)}(0) (-1)^{k-r} \right]$$

$$= - \{ x^{2p+2} / [(p+1)!]^2 \} \left[ \sum_{p+1}^{-1} \sum_{p+1}^{-1} - c^{(2p+2)}(0) (-1)^{p+1} \right].$$

; this last expression is positive since  $c^{(2p+2)}(0) (-1)^{p+1} = \sum_{p+1, p+1}$ .

Thus under the added assumption that  $c(\cdot)$  has derivatives up to the order of  $4p+4$ ,  $\text{Var}(\eta_i^*) = \text{Var}(\eta_i) = \delta^2 |t_i - t_{n+1}|^{2p+2}$  for some constant  $\delta^2$ ; the two models would then agree locally.

In our analysis we will adopt the structural model with  $\text{Var}(\eta_i) = \delta^2 |t_i - t_{n+1}|^{2p+2}$ . It is the simpler of the two. They agree at least locally, as was pointed out above and it is the local behavior which is of primary importance. As  $|t_i - t_{n+1}|$  increases,  $\eta_i$  becomes dominant,  $R_i$  reduces to noise and is essentially "windowed out" of inferences about  $\beta$ . Of course, the proposed model would be more appealing in principle if  $\text{Var}(\eta_i)$  were a bounded function,  $h$ , of  $|t_i - t_{n+1}|$ . However, in practice, this would appear to make almost no difference in estimating  $\beta_0 = S(t_{n+1})$ . A variety of  $h$ 's, bounded and unbounded, were tried experimentally on the data in the application presented in Section 3 with little effect on the outcome. The assumed underlying joint distribution was normal throughout that analysis.

2.5 STATISTICAL INFERENCE. Conditionally for fixed  $\beta$  assume that equation (2.1) holds with  $\text{Cov}(c) = \sigma^2 \text{diag} [1+c x_1^{2p+2}, \dots, 1+c x_n^{2p+2}] = \sigma^2 H$ , say where  $\sigma^2 = \text{Var}(\eta_i)$  is the variance of the noise,  $c = \delta^2/\sigma^2$ , and  $\delta^2$  represents uncertainty about the size of  $D^{p+1}S/(p+1)!$  in the remainder term of the Taylor expansion of  $S$  and  $H = \text{diag} [h_1, \dots, h_n]$ ,  $h_i = 1+c x_i^{2p+2}$  for all  $i$ . Assume all a priori unspecified quantities have a jointly normal distribution. The following results are standard (see Lindley and Smith 1972) with " $\sim$ " meaning distributed as and " $U|V$ " meaning the conditional distribution of  $U$  given  $V$ :

$$R|\Sigma, \sigma^2, H \sim N(X\beta^0, X[\Sigma^{-1} + \sigma^2 H]) \quad (2.5)$$

$$\beta|R, \Sigma, \sigma^2, H \sim N(F[\Sigma^{-1}\beta^0 + X^T H^{-1}R], F)$$

where  $F = (\Sigma^{-1} + \sigma^{-2} X^T H^{-1} X)^{-1}$ .

In principle this solves the inference problem. When the hyperparameters in  $\beta^0$ ,  $\Sigma$ ,  $\sigma^2$  and  $H$  are specified,  $\beta$  may be estimated by its a posteriori expectation and a credibility ellipsoid is easily constructed to give a 95% credibility set for  $\beta$ . Alternately, point and credibility interval estimators of  $\beta_0 = S(t_{n+1})$ ,  $\beta_1 = DS(t_{n+1})$ , etc., may be constructed from the normal marginal distributions of these parameters.

Usually, additional stages would need to be added to the process of constructing the prior since specifying the hyperparameters at the first stage would be difficult. Moreover, the resulting theory would be technically complex.

A somewhat ad hoc, much simpler, approach is adopted for the analyses of Section 3. Observe that as  $\Sigma^{-1} \rightarrow 0$ ,  $F \rightarrow \sigma^2 (X^T H^{-1} X)^{-1}$  so that  $E(\beta|R) \rightarrow (X^T H^{-1} X)^{-1} X^T H^{-1} R$ , the generalized least squares estimator of  $\beta$  with (a posteriori) covariance matrix  $\sigma^2 (X^T H^{-1} X)^{-1}$ . Thus, using a diffuse prior for  $\beta$  leads to the classical estimator when  $\sigma^2$  and  $H$  are specified.

The same estimator may be obtained by treating  $\beta$  as a hyperparameter in the likelihood obtained from the conditional density of  $R$  given  $\beta$ ,  $\sigma^2$ , and  $H$ ; evaluated at  $R = r$ , this is, ignoring irrelevant constants,

$$\sigma^{-n/2} |H|^{-1/2} \exp \left\{ -(\sigma^{-2}/2) (r-X\beta)^T H^{-1} (r-X\beta) \right\} .$$

By standard theory this is maximized at

$$\begin{aligned} \hat{\beta} &= (X^T H^{-1} X)^{-1} X^T H^{-1} r \\ \hat{\sigma}^2 &= n^{-1} (r-X\hat{\beta})^T H^{-1} (r-X\hat{\beta}) . \end{aligned} \quad (2.6)$$

And a priori, conditionally, given  $\beta$  and  $H$ ,  $\hat{\beta} \sim N(\beta, \sigma^2 (X^T H^{-1} X)^{-1})$ , independently of  $\hat{\sigma}^2$  which has a chi-squared distribution with  $n-p-1$  degrees of freedom.

Only  $c$  or equivalently  $\delta^2$  remains to be chosen. If  $p = 1$  or  $2$ , say, this may be done by introspection. Alternately since the likelihood, with  $\sigma^2 = \hat{\sigma}^2$  and  $\beta = \hat{\beta}$ , reduces to

$$(\hat{\sigma}^2)^{-n} |H|^{-1/2} , \quad (2.7)$$

an estimate  $\hat{c}$  may be found by maximizing, numerically, the expression displayed in (2.7). The stochastic independence of  $\hat{\beta}$  and  $\hat{\sigma}^2$  is then lost and the resulting credibility interval for  $\beta_0 = S(t_{n+1})$  is then only approximate.

In the application of Section 3,  $c = 0$  was obtained for both  $p=1$  and  $2$  by maximizing equation (2.7) but the reduced likelihood is nearly flat in a large neighbourhood of  $c=0$  so this choice is not strongly supported. Introspection suggests a small value for  $\delta^2$  and hence  $c$ , but  $c=0$  seems unrealistic. The possible dependence of the maximum likelihood estimate of  $c$  on the choice of  $t_{n+1}$  presents a dilemma. A data-based alternative to maximum likelihood for finding  $c$ , which avoids this issue, was finally adopted for the application of Section 3. This approach uses cross-validation (Stone 1974);  $t_{n+1} = t_i$  is chosen, successively,  $r_i$  is dropped from the sample and interpolated using  $\hat{c}_1 = \hat{\beta}_0(i)$  obtained as above by maximum likelihood, and then  $c = c$  is found by numerically minimizing  $\sum (r_i - \hat{c}_1)^2$ . This approach gave a small but positive

value of  $c$  and was judged superior to the maximum likelihood method. It is noteworthy that  $c$ , found in this manner, does not depend on  $\sigma^2$ .

To estimate  $\sigma^2$  observe that for  $t_{n+1} = t_1$ , the a priori local structural model is  $R_i = \alpha_i + \sigma Z_i$   $i = 1, \dots, n$  where  $\alpha_i = S(t_i)$  and  $Z_i$  has mean 0, variance 1. Now  $\alpha_i$  may be estimated without knowledge of  $\sigma$  by the generalized least squares method described above and  $\text{Var}(\hat{\alpha}_i) = c_i \sigma^2$  where, given  $c = \hat{c}$ ,  $c_i$  is a known factor obtained from the generalized least squares procedure itself. Thus  $E(R_i - \hat{\alpha}_i)^2 = \text{Var}(\hat{\alpha}_i)^2 + \sigma^2 = \sigma^2(1 + c_i)$  when the noise is regarded as independent of the carrier. Consequently an unbiased estimated estimator of  $\sigma^2$  is

$$\hat{\sigma}^2 = \frac{1}{n} \sum (R_i - \hat{\alpha}_i)^2 / (1 + c_i).$$

This is the estimator which is used in the application given in the next section.

The proposed estimator  $\hat{\beta}$ , of  $\beta$  has the curious property that for large

$$|t_{n+1}| \quad \hat{\beta}_r = \hat{\lambda}_r + \hat{\lambda}_{r+1} t_{n+1} + \dots + \hat{\lambda}_p t_{n+1}^{p-r} / (p-r)! \quad , \quad r = 0, 1, \dots, p \quad (2.7)$$

where the  $\lambda$ 's are the least squares estimates obtained by minimizing  $\sum_i (r_i - \lambda_0 - \lambda_1 t_i - \dots - \lambda_p t_i^p / p!)^2$ . To see this observe that  $\hat{\beta}$  is found by minimizing

$$\sum_i (r_i - \beta_0 - \beta_1 (t_i - t_{n+1}) - \dots - \beta_p (t_i - t_{n+1})^p / p!)^2 h_i^{-1}$$

where  $d_i = 1 + c(t_i - t_{n+1})^{2p+2} = c|t_{n+1}|^{2p+2} = d_j$ ,  $i, j=1, \dots, n$  when  $|t_{n+1}|$  is large. Thus  $h_i^{-1}$  may be replaced by 1 in the weighted residual sum of squares which determines  $\hat{\beta}$ . Now let  $\beta_r = \sum_{j=r}^p \lambda_j t_{n+1}^{j-r} / (j-r)!$  where the  $\lambda$ 's are arbitrary,  $r = 0, 1, \dots, p$ . Then  $\sum_r \beta_r (t_i - t_{n+1})^r / r! = \sum_j (\lambda_j / j!) \sum_r \binom{j}{r} t_{n+1}^{j-r} (t_i - t_{n+1})^r$  after interchanging the orders of summation and this in turn is just  $\sum_j \lambda_j t_i^j / j!$  which proves the assertion.

3. APPLICATION. In this section, the results of the last are tested on a specific data set. These data are derived from the ADS data-base which is maintained and updated by Battelle's Pacific Northwest Laboratories in Richland, Washington. The data were obtained from one of the nine stations which constitute the MAPSS/PCW, monitoring network. This station, located at Pennsylvania State University, has latitude and longitude  $40^{\circ}47'18$  and  $75^{\circ}56'47$  respectively. This network is found in the Northeastern part of the United States and has provided event-based chemical measurements of wet deposition events since about 1976, although the start-up varies from station to station. For illustrative purposes, our analyses are confined to two subcases: (i) the field pH values which were measured at a particular station during the precipitation events of 1977, and (ii) the monthly average pH-values over the 1976-82 period at the same station. The last is an intrinsically smoother alternative to the former and will be referred to as the Monthly Average Data. Throughout this section time is measured in days starting with January 1, 1976 which is regarded as the origin.

Only locally constant, linear and quadratic (i.e.  $p \leq 2$ ) structural models will be considered in this illustrative application. In the locally linear case, for example, with  $w_i = t_i - t_{n+1}$  the a priori model,  $R_i = \alpha + \beta w_i + \sigma(1+cw_i^2)^{1/2} Z_i$ ,  $i=1, \dots, n$ , obtains, given  $\alpha$ ,  $\beta$ ,  $\sigma$ , and  $c$ , where  $\alpha = S(t_{n+1})$  is the interpolant or predictand, that is, the object of primary inferential interest,  $\beta = S'(t_{n+1})$ , and  $Z_i \sim N(0,1)$  independently of each other and of  $\alpha$ ,  $\beta$ ,  $c$ , and  $\sigma^2$ . Under appropriately diffuse prior distributions, posterior estimates of  $\alpha$  and  $\beta$ , may be found by maximum likelihood when  $c$  and  $\sigma^2$  are fixed. The results are precisely those obtained by generalized least

squares. Thus the calculations are easily performed by using conventional statistical packages like SAS or by simple-to-write code (which in our case was 1/30 the cost of the first approach). The computations for the locally constant and locally quadratic cases are just as straightforward as those for the locally linear case; the 4-th power of  $w_1$ 's in  $R_1$ 's conditional variance (given  $a$  and  $B$ ) is replaced in the former and latter cases, respectively, by 2nd and 6th power, according to the reasoning of Section 2. The estimated value of  $c$  was found by cross-validation using a numerical optimization subroutine to minimize the sum of the leave-one-out squared errors as described in Section 2.5.  $\hat{\sigma}^2$  was also found by the method described in that section. The estimated variance of  $a$  is easily found by means of standard formulas from the theory of curvilinear regression. Since a posteriori,  $a$  has a normal distribution with mean  $a$ , an approximate 95% credibility interval for  $a$  is obtained by adding to and subtracting from  $\hat{a}$  twice the estimated standard deviation of  $a$ .

Before turning to an analysis of the acid deposition data, these methods were applied to an artificial, parabolic data set. The locally quadratic model gave a perfect fit with estimated  $\sigma^2$  and  $c$  values of 0. Moreover, the locally quadratic extrapolant of these data is the same parabola as that which generated the data. This is because the locally quadratic fit is also the least squares quadratic fit and, as pointed out in Section 2 in the material following equation (2.7), the extrapolant of any of the proposed local fits is the least squares polynomial fit of the same order, extended beyond the data. In Figure 1, the locally constant and linear fits are displayed. The local fits are again essentially perfect over the range of the fake data. However, the extrapolants are asymptotically constant, in the locally linear case

because the least squares line for these data has slope 0.

FIGURE 1 HERE

Observe in Figure 1 how much more abruptly than the locally linear fit, the locally constant fit, which is not required to be locally differentiable, turns toward its right hand asymptote, the sample average of r-values.

The successively greater smoothnesses of the locally constant, linear, and quadratic-fits are revealed in Figure 2 where all three models are fitted to the 1977 data. The twice differentiable quadratic cannot make the quick turns of the linear, for example, and it tends to round off the latter's corners. Such comparisons are somewhat tentative, however, since the "smoothing parameter",  $c$ , is fitted independently for the three curves and its size determines to a considerable extent the smoothness of the curve. When  $c$  is large, the conditional a priori variance of  $R_i$  grows very rapidly as  $|w_i| = |t_i - t_{n+1}|$  increases, the fit at any given  $t_{n+1}$  is then highly localized and only the  $R_i$ 's with  $|w_i|$  close to zero determine that fit. The result for a large  $c$  is very irregular for a noisy process like that portrayed in Figure 2. As a final note about Figure 2, observe that because of a degree of some apparent seasonality in the data they have a slightly parabolic character. Thus the locally quadratic fit, unlike its linear counterpart emerges from the data at either extremity (side) with a marked upward trend. This emphasizes the need to remove seasonal components, if any, by adding a model to the smooth to avoid the possibility of nonsensical predictions.

FIGURE 2 HERE

The same caution is necessary in interpreting the credibility intervals portrayed in Figure 3. It must also be emphasized that just as the results in Figure 2 are not curve estimates, the bands in Figure 3 are not simultaneous interval estimates. That is they cannot be said to enclose  $\{S(t), a < t < b\}$ , over the appropriate time interval  $(a,b)$ , with probability 95% even approximately. Such bands would be appreciably wider. Those in Figure 3 merely indicate the appropriate credibility interval for  $S(t_{n+1})$  at each  $t_{n+1}$ . These are not tolerance intervals, that is, they do not contain  $R(t_{n+1})$  with probability 95%. Such intervals are easily obtained approximately, however, by adding (respectively subtracting)  $2\sigma$  to the upper (respectively lower) limit of  $S$ 's credibility intervals. Although it would seem to be of little interest we did examine the locally constant case with monthly data and found that 5 of 80 i.e. 6% of  $R$ -values were not covered in the appropriate tolerance intervals; in that case  $\hat{\sigma} = 0.1526$ .

FIGURE 3 HERE

With these cautionary remarks in mind let us turn, to Figure 3 for an evaluation of the results. The credibility band for the locally quadratic model turns out to be narrower than that of the locally linear model. This is intuitively appealing; the smoothing constant  $c$  is a critical determinant of the width of these bands and represents essentially the size of the remainder term in the Taylor expansions. It is intuitive that this should be smaller in

the quadratic case since there, the second order term has been "lifted" out of the error for the linear case and so reduced it. The cross validatory estimates of  $c$  agree with this choice in the examples, but there is no guarantee that a similar result would obtain in every case. Notice that the credibility intervals for the two cases represented in Figure 3, are considerably different outside the range of the data; the quadratic band is much narrower.

Figure 4 reveals that the locally linear bands are somewhat anomalous outside the range of the data. They diverge more rapidly than either those of the locally constant or locally quadratic case. The apparent anomaly is a product of the ad hoc method used for fitting  $c$ ; the cross validatory choice in the locally linear case is too large relative to the other two cases.

FIGURE 4 HERE

The residuals for the locally linear fit for both data sets are represented in Figure 5. The assumption of a constant error variance,  $\sigma^2$ , seems reasonable in the case of the 1977 data. However, the residuals for the monthly average data make this hypothesis seem more tenuous in that case. Overall the fits seem reasonably good.

FIGURE 5 HERE

In Figure 6 are contrasted the results of linear and locally linear fitting when the middle of the data are deleted. Notice in particular that

the width of the confidence band for the linear fit reaches a minimum where the data are missing. This is, technically, because the time mean,  $t$ , is located in the gap. However, this is somewhat paradoxical since there would seem to be considerable uncertainty in this region because of the lack of data. The explanation lies in the unequivocal choice of the linear model over the whole range of the data. In contrast, the width of the credibility interval for the locally linear fit increases in the gap as would be expected.

FIGURE 6 HERE

In Figure 7 we illustrate the curve of estimates of  $DS(t)$  for varying values of  $t$  over the 1977. The four peaks indicate the times of steepest ascent of the pH smooth and the troughs, the times of steepest descent. Credibility intervals could well have been plotted here; but we declined to do so as no new issues arise.

FIGURE 7 HERE

Finally in Figure 8 are displayed a number of locally linear fits with varying rather than fitted values of  $c$ . This illustrates the great range of possible smoothers which can be achieved under the locally linear assumption alone. The choice  $c = 0$  gives the least squares line, while for large  $c$ , the fit tends to follow very closely the individual data points.

FIGURE 8 HERE

4. RELATED WORK. There is a vast literature on the subject of this paper and we will not attempt anything like a complete survey. Our review will cite just the most relevant, recent work to help put our results into perspective.

Our results intersect with those of Cleveland (1979) who presents a generalized least squares, locally polynomial, two dimensional scatterplot smoother. For each  $t_{n+1} = t_1, \hat{\lambda}_0, \hat{\lambda}_1, \dots, \hat{\lambda}_p$  are chosen to minimize  $\sum w_1(t_j) (\tau_j - \lambda_0 - \lambda_1 t_j - \dots - \lambda_p t_j^p)^2$  as a function of the  $\lambda$ 's. Thus if  $w_1(t_j) = [1 + c (t_j - t_1)^{2p+2}]^{-1}$ , the smoothed value of  $\tau_1, \hat{\tau}_1$  say, obtained by Cleveland will equal the estimate proposed in this paper for  $S(t_1)$ . Cleveland does not, however, suggest this weight function because his analyses, unlike ours, is not driven by an underlying model and its Taylor expansion. His weight functions,  $w_1(t_j) = [(1 - |t_j - t_1|^2)_+^2]^{-1}$ , for example, derive instead from exploratory data analytic and robustness considerations. In fact, Cleveland's smooth is obtained iteratively by a succession of minimizations where at each step, the weight function changes according to the character of the residuals from the previous step. His final smoother is found by connecting the  $\hat{\tau}_1$  by line segments whereas we fit an estimate of  $S(t_{n+1})$  at each point,  $t_{n+1}$ . Our work differs from Cleveland's in other ways. Our theory is Bayesian and, in this paper, empirical Bayesian; in general our smoother would not be a generalized Bayes procedure. We, unlike, Cleveland, are concerned with extrapolation outside the data set and with reliability bounds on our extrapolated and interpolated values. And most importantly, our theory has an conceptually easy extension to the space-time context.

We share with Fearn (1975) a multivariate normal based Bayesian approach to the analysis of individual sample paths, in his case, growth curves. But our paths diverge when he adopts a global linear model for the carrier. For individual  $i$ , the observable vector  $R_i$  has a conditional a priori distribution,  $R_i | \beta_i, \sigma^2, \sim N(X_i \beta_i, \sigma^2 I)$ ,  $i = 1, \dots, m$ . And  $\beta_i | \mu, C \sim N(\mu, C)$  for all  $i$ . An ad-hoc procedure is used to estimate the  $\sigma^2$  and  $C$ .

Because we do not postulate a global, parametric model, our approach to modeling offers far greater flexibility than Fearn's while retaining all the advantages of the Bayesian approach. Similar gains in flexibility are achieved over Hui and Berger (1983) who also assume underlying normality in their approach which assumes  $R(t, \omega_i) = a_i + b_i t$ , for  $t$  in the (short) follow-up interval for individual  $i=1, \dots, n$ . Their analysis is not as local as ours and their results are quite different in character from those presented here. They do consider the problem of estimating the unspecified covariance matrix.

O'Hagan (1978) concerns himself with local regression curve fitting and design. His most general model has  $R = R(x) \sim N(\eta(x), \xi(x))$ . The correlation of  $R(x)$  and  $R(x^*)$  is through the correlation of  $\eta(x)$  and  $\eta(x^*)$  which are assumed to have jointly, a normal distribution, as are any finite set of  $\eta$ 's. The posterior mean of  $\eta(x)$  is readily found and is the posterior mean of  $R(x)$ , which is thus readily interpolated or predicted as appropriate. O'Hagan (ibid) regards  $\xi(\cdot)$  as specified but the case where it is not is briefly discussed. His work is similar in spirit to that presented here. It differs in as much as he does not exploit as we do, the natural, local structural model which regularity entails and which forces  $S(t)$  and  $S(t_{n+1})$  into a linear relationship (with error). We view our methodology as an ineluctable consequence of our regularity assumption and

therefore necessarily favor it over O'Hagans'. We favor our approach too because our structural model guides the quantification of prior opinion by prescribing the framework on which it must be attached. It would therefore be simpler to apply in practice.

The considerable literature on spline smoothing techniques is well surveyed by Silverman (1985) who gives an extensive bibliography. The polynomial smoothing spline of order  $2m-1$  is found by minimizing, with respect to  $S$ , a positive multiple of  $\sum (r_i - S(t_i))^2 + \lambda \int [D^m S(t)]^2 dt$ ; this integral is over  $[0,1]$  in Wahba (1983) where it is assumed that  $R(t_i) = S(t_i) + \epsilon_i$  and that  $\epsilon_i$  are zero mean normal disturbances of constant variance for all  $i$ . The result is a piecewise polynomial of order  $2m-1$  in each interval  $(t_i, t_{i+1})$  with smooth joins at the interval boundaries for all  $i$ . Wahba (1983) points out and exploits a Bayesian interpretation of the smoothing spline, that is a Bayes estimate with respect to a certain zero mean Gaussian prior. And she is able to develop intervals for the  $S(t_i)$  which seem to have the confidence property, according to a simulation study she describes in her paper (with  $\lambda$  estimated by cross validation). It should be emphasized that these are not simultaneous confidence bands, i.e. the percentage of the  $S(t_i)$  which would lie in their intervals would average around 95%. It is not the case that all of the  $S(t_i)$  would simultaneously be covered around 95% of the time. It is not clear if simultaneous confidence intervals would be achievable with Wahba's theory. Presumably intervals (credibility or confidence) for  $S(t_{n+1})$ , an interpolant or predictand, could be obtained but, even though this is a problem of considerable practical significance, it is not addressed by Wahba.

The spirit of spline-smoothing lies in frequency theory so it is not

altogether meaningful to try to compare and contrast such work with ours. However, a few comments would seem in order. The Bayesianity of spline smoothers is an artifact, a product, in particular, of the objective function which is minimized in their construction. It would seem that the derived prior is essentially unique so that the applicability of these smoothers as genuinely Bayesian procedures would seem to be severely limited. Beyond this, the prior, a distribution for the infinite dimensional "parameter",  $S(\cdot)$ , is an overspecification in the sense that when inference is about  $S(t_{n+1})$ , for example, only a distribution with a finite dimensional support set is required. A major conceptual disadvantage of spline smoothers pointed out by Silverman is that they are defined only implicitly as the solution to a minimization problem except for large samples when an approximate explicit expression may be derived. In contrast, our method does yield an explicit estimate of  $S(t_{n+1})$  along with explicit credibility intervals for this estimate. And although our extensions of our present work are incomplete, it does seem to promise easy analogues of the present work in the case of vector-valued response functions of a vector argument. In contrast the theory of spline-smoothers seems to present considerable technical difficulty and while the work of Wabba (1983) has been extended, these extensions are hard-won and quite limited to date.

There has been a considerable interest in nonparametric regression and the celebrated paper of Stone (1977), among others, discusses the consistency of the members of classes of such procedures. There is little intersection with our work which is more concerned with generating non-parametric regression function estimators than on repeated sampling properties of the

results. Stone (1977) does consider in passing a locally linear estimator and the same estimator appears more informally in Chambers, Cleveland, Kleiner and Tukey (1983, p.96). The locally linear estimators in both of the just cited works differs from that which emerges here in the locally linear special case; our estimator is not, in fact, linear in the regressor,  $t$ .

A notable, recent contribution to the literature of smoothing is the paper of Breiman and Friedman (1985). Their very general approach yields an implicit characterization of transformations of the regressand and regressor, which maximizes the correlation between the transformed variables. Their paper is in a nonparametric frequency setting and has little intersection with ours. In particular they do not consider the problem of developing reliability bands or the prediction problem. Rather their work is concerned with finding a suitable algorithm for determining the required optimal transformations.

5. DISCUSSION. This report presents a very general approach to the analysis of stochastic processes,  $R = C + N$  where  $C = M + S$ ,  $M$  is a parametric model (possibly zero),  $S$  is a smooth function which is at least continuous and possibly differentiable to some order and  $N$  is uncorrelated noise. The method is flexible and easy to implement. In its generality it subsumes space-time series and the regression problem. It allows for the analysis of time series with only a single observation per sample path and in general gives explicit answers which are much different in form to those derived by classical methods.

The response  $R$ , would in general be a matrix whose rows (as few as one, possibly) would correspond to independent replicates and columns to possibly correlated, per subject responses. On each row would be stacked the response vectors, one for each combination of quantitative levels at which the responses were measured. These levels could include such things as space-time co-ordinates and independent (continuous) variables like, say temperature.

The model  $M$  incorporates all available prior knowledge about  $C$ 's functional form. So  $M$  could well contain terms for trend and seasonality. And  $M$  should incorporate all of  $C$ 's discontinuous components. For example,  $c(t) = \theta$  or  $\theta$  according as  $t \leq t_0$ , or  $t > t_0$ , with  $\theta$  to be fitted, would incorporate the effect of an intervention at time  $t_0$ . Such an intervention might be the closure of a smelter or alternatively the start-up of a new power generating facility; in either of these cases,  $R$  might well be the pH of wet acidic deposition. In any case,  $M$  would be analyzed in a conventional fashion and in this paper we suppose  $M \equiv 0$ .

$S$  is the focus of this paper. The responses,  $R$ , are assumed to have been measured at distinct levels  $t = t_1, \dots, t_n$ , of a single, continuous explanatory factor. And inference is taken to be about  $S(t_{n+1})$ , at a point,  $t = t_{n+1}$ , where, possibly, no observations have been taken. The relationship between  $S(t_i)$ ,  $i=1, \dots, n$  and  $S(t_{n+1})$  is dictated by Taylor's theorem;  $S(t_i) = S(t_{n+1}) + \dots + \text{remainder}$ . This relationship is an inevitable consequence of the assumed regularity, it is exact, and it provides the vital link for inference between the  $R(t_i)$ ,  $i=1, \dots, n$ , and  $S(t_{n+1})$ . A priori uncertainty about the remainder and about all other such unspecified items is quantified in terms of probabilities. The remainder is not independent of the other terms in the expansion. However, we show in Section 2, that this assumption is approximately valid for  $t_i$  near  $t_{n+1}$ , and this is all that matters because the more remote observed  $R$ -values are windowed out. The Taylor based structural model reduces enormously the need for introspection in a priori distribution elicitation.

The use of Bayesian methods has the advantage that prior knowledge about a priori unspecified parameters can be accommodated in the analysis to a far greater degree than merely through the choice of  $M$ . On the other hand by adopting appropriate diffuse prior probability distribution, essentially empirically Bayes procedures like those which form the core of results in this paper, are obtained. Empirical Bayes methods are enjoying a good deal of current popularity because of their capability to transfer information between subjects ("borrowing from strength") and often, thereby, improving the overall quality of the analysis.

Our approach has the advantage that estimates of  $S(t_{n+1})$  are obtained in an explicit form. They are linear functions of the  $R(t_1)$  and a combination of rational functions of  $t_{n+1}$ . For extreme values of  $t_{n+1}$ ,  $S(t_{n+1})$  is, in the case considered here where the priors are vague, predicted by the least squares polynomial fitted to the data; its order is that of the highest term of the Taylor expansion. However, in a proper Bayesian analysis of this problem more controlled predictions would be obtained.

In any case, credibility intervals for  $S(t_{n+1})$  are easily computed in explicit form from its a posteriori distribution. And in the case considered here, this interval should have, approximately, the confidence property. Simultaneous bands are desirable in principle for  $(S(t_{n+1}), \dots, S(t_{n+k}))$ ,  $k \geq 1$ .

Our approach readily yields quantiles for  $S(t_{n+1})$  along with its mean value. And it can be applied to make inferences for differences like  $S(t_{n+1}) - S(t_{n+2})$ .

It should be emphasized that although S-curves and credibility bands are pictured in the Figures these are intended to be applied on a point-by-point basis and not simultaneously. Simultaneous credibility bands would be a lot wider.

It should also be emphasized that the distributions in this paper have a finite dimensional support generated by the quantities emerging at the times  $t = t_1, \dots, t_{n+1}$  and these distributions are thought of as just approximate

expressions of uncertainty. We have not considered the technically challenging problem of studying the class of stochastic processes,  $\{S(t): 0 \leq t \leq \infty\}$ , for which our local structural or stochastic linear models hold. An understanding of these processes would be helpful in insuring coherence in elicited prior probabilities.

#### REFERENCES

- Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation (with discussion). Jour. Amer. Statist. Assoc., 80, 580-597.
- Chambers, J.W., Cleveland, W.S., Kleiner, B., Tukey, P.A. (1983). Graphical methods for data analysis. Belmont, California: Wadsworth International Group.
- Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. Jour. Amer. Statist. Assoc., 74, 829-836.
- Fern, T. (1975). A Bayesian approach to growth curves. Biometrika, 62, 89-100.
- Hui, S.L. and Berger, J.O. (1983). Empirical Bayes estimation of rates in longitudinal studies. Jour. Amer. Statist. Assoc., 78, 753-759.
- Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). Jour. Roy. Statist. Soc., Ser. B, 34, 1-41.
- Macaulay, Frederick, R. (1931). The smoothing of time series. New York: National Bureau of Economic Research.
- O'Hagan, A. (1978). Curve fitting and optimal design for prediction (with Discussion). Jour. Roy. Statist. Soc., Ser. B, 40, 1-42.
- Priestly, M.B. (1981). Spectral analysis and time series. London: Academic Press Inc.
- Silverman, B.W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). Jour. Roy. Statist. Soc., Ser. B, 47, 1-52.
- Stone, C. (1977). Consistent non-parametric regression, Ann. Statist., 5, 595-645.
- Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. Jour. Roy. Statist. Soc., Ser. B, 36, 111-147.
- Wahba, G. (1983). Bayesian "confidence intervals for the cross validating smoothing spline" Jour. Roy. Statist. Soc., Ser. B, 45, 133-150.

FIGURE 1. Locally Constant (broken line) and Linear (solid line) Fits to Artificial (parabolic) Data.

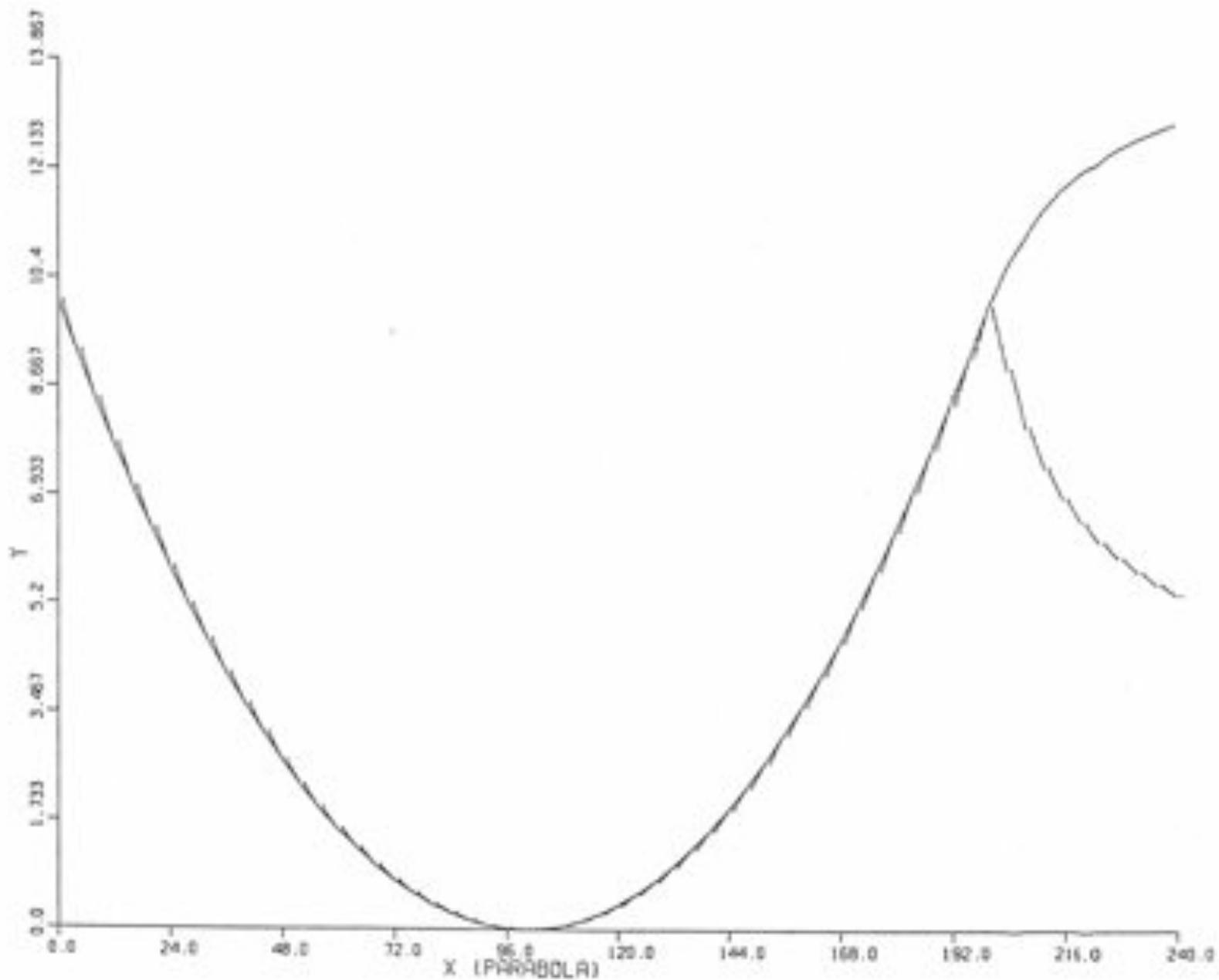


FIGURE 2. A Comparison of the Locally Constant (dotted line), Linear (broken line) and Quadratic (solid line) Fits to 1977 pH Data.

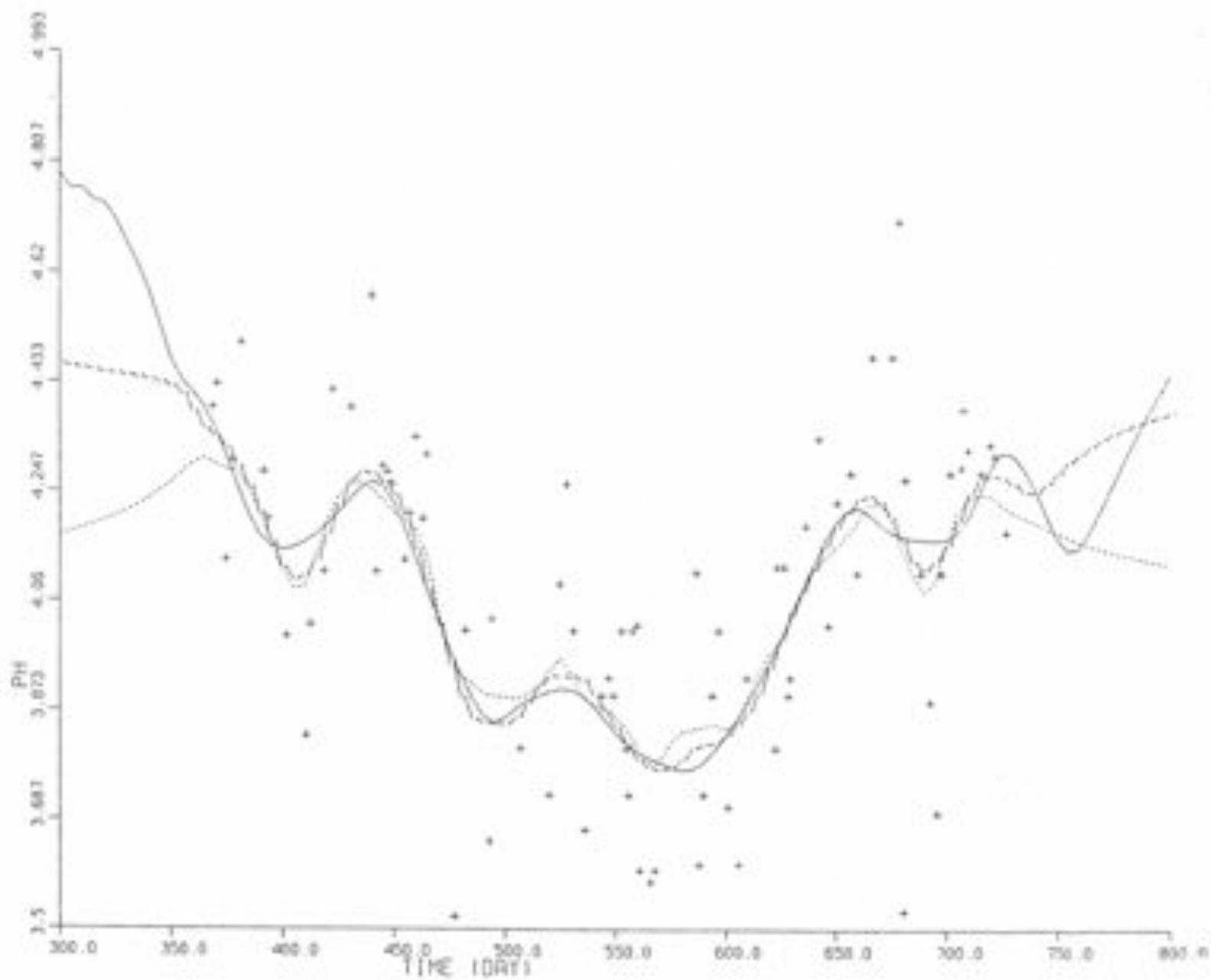


FIGURE 3. A Comparison of the Credibility intervals for Locally Linear (broken line) and Quadratic (solid line) Fits for the 1977 (top) and Monthly Average (bottom) pH Data.

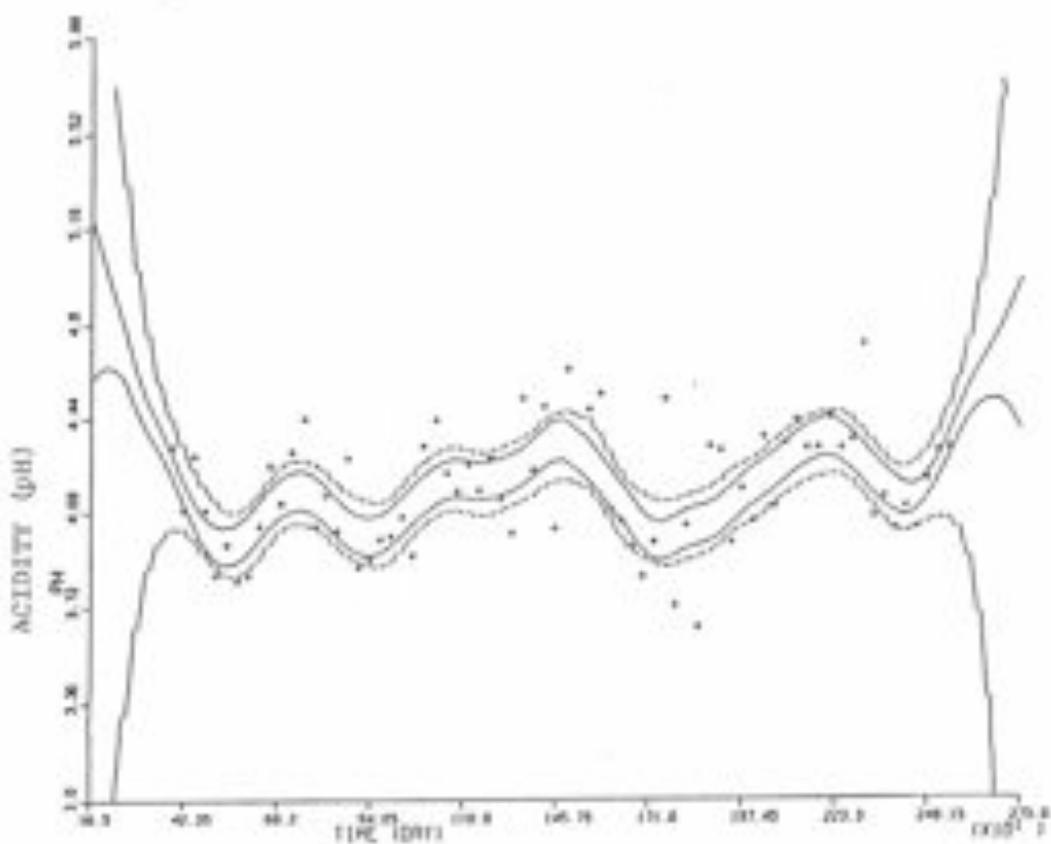
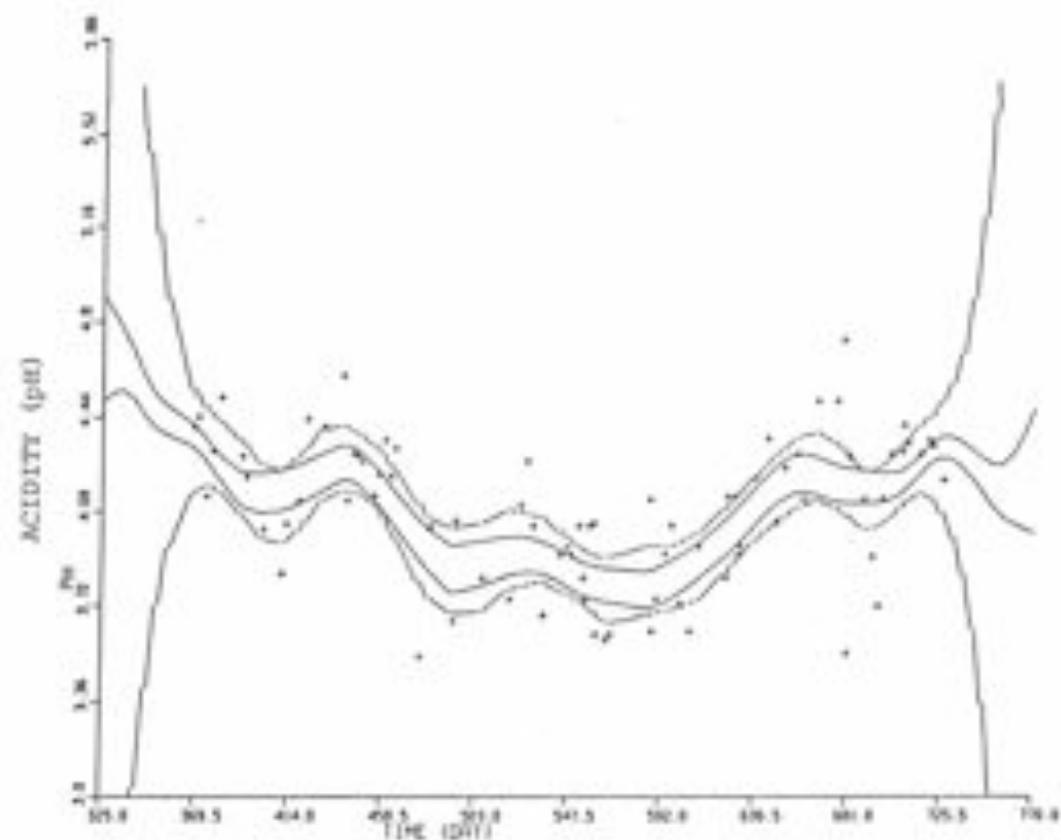


FIGURE 4. Locally Constant, Linear, and Quadratic Fits for 1977 and the Monthly Average pH Data.

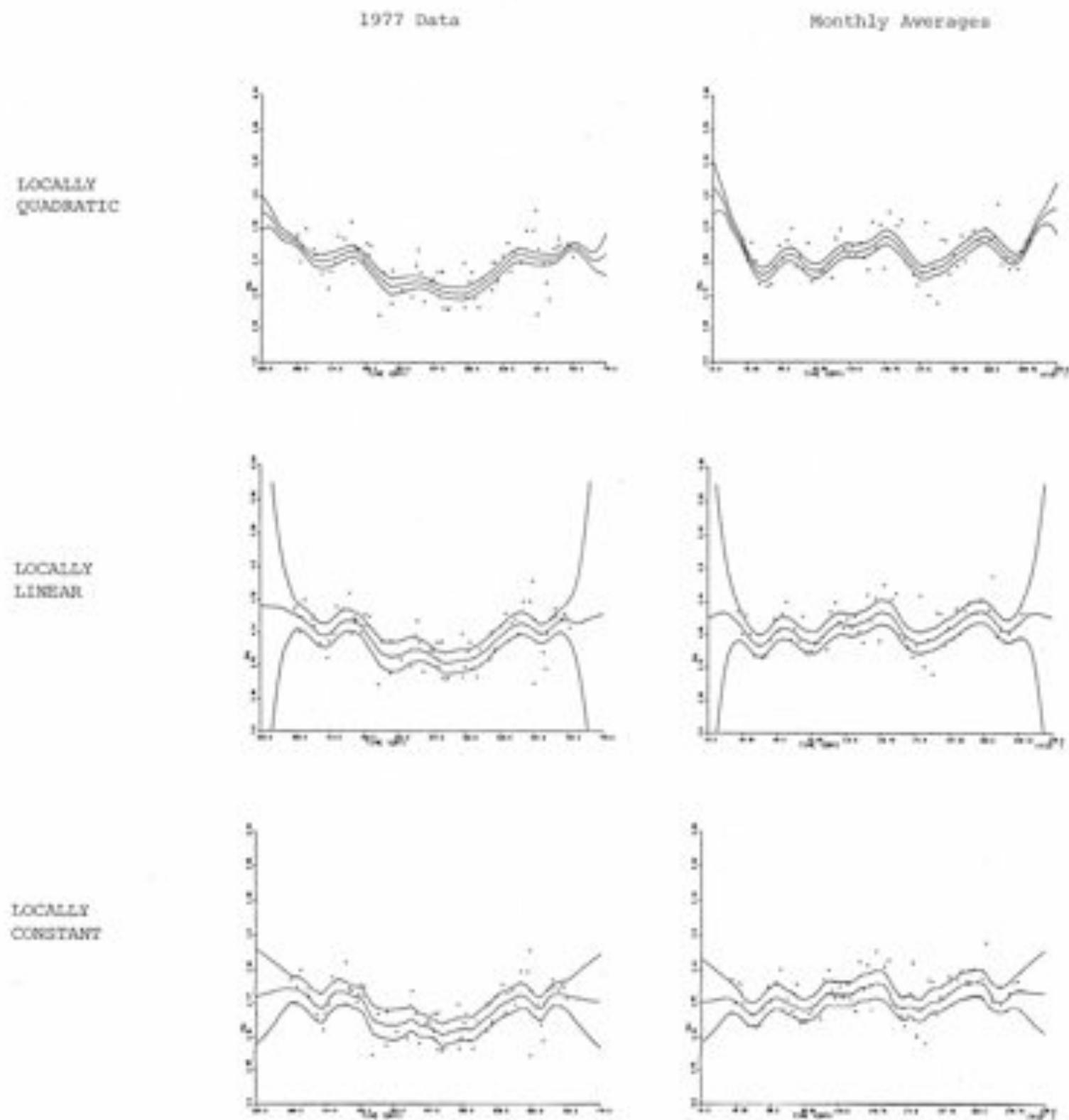


FIGURE 5. Residual Plots from the locally Linear Fits to the 1977 (top) and the Monthly Average (bottom) pH Data.

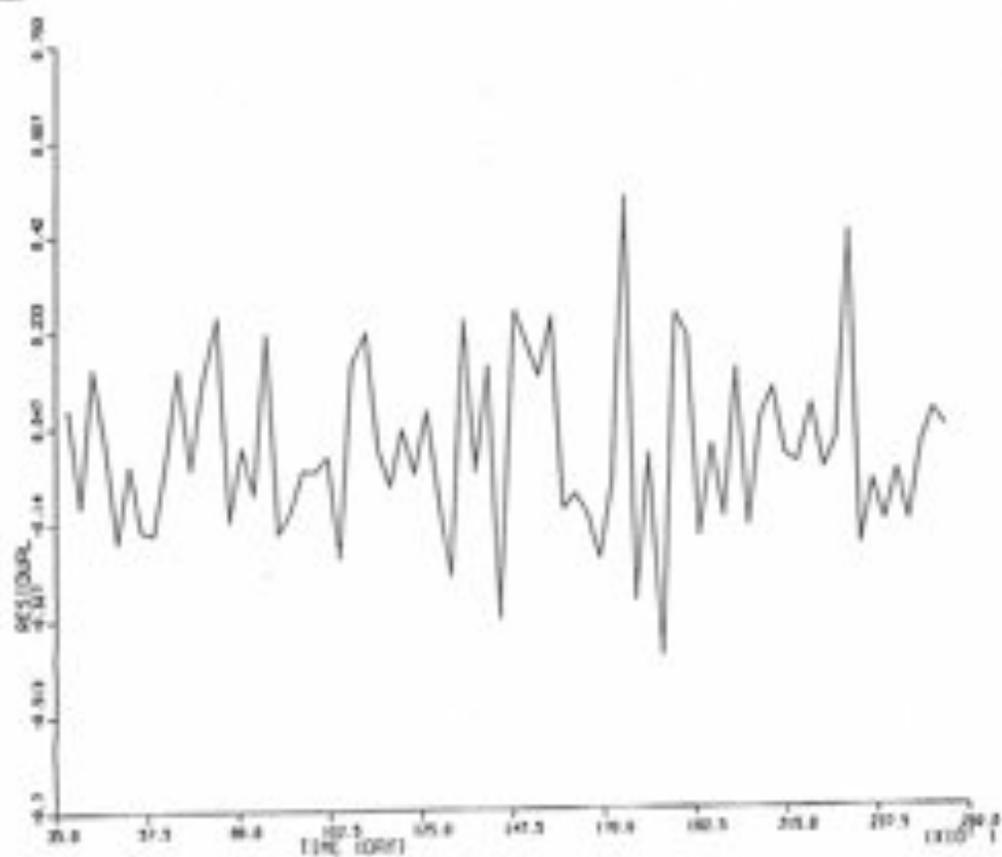
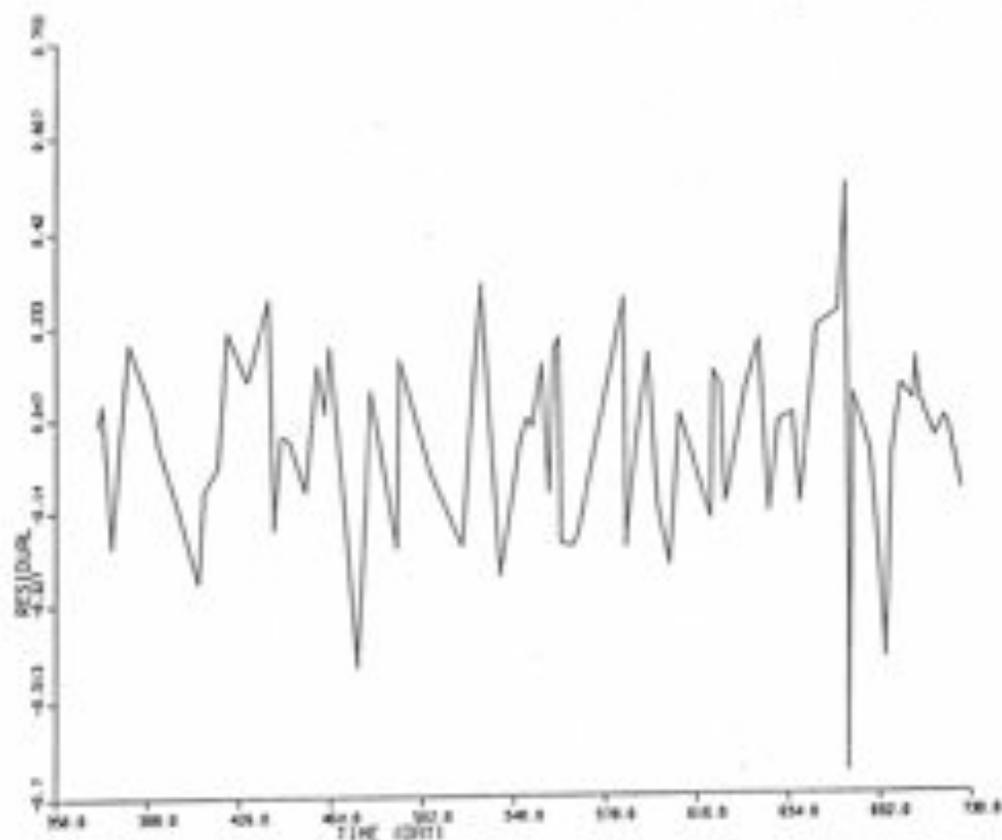


FIGURE 6. Comparison of the Credibility band for the Locally Linear Fit and the Confidence Band for Linear Regression: Central Data Points Deleted.

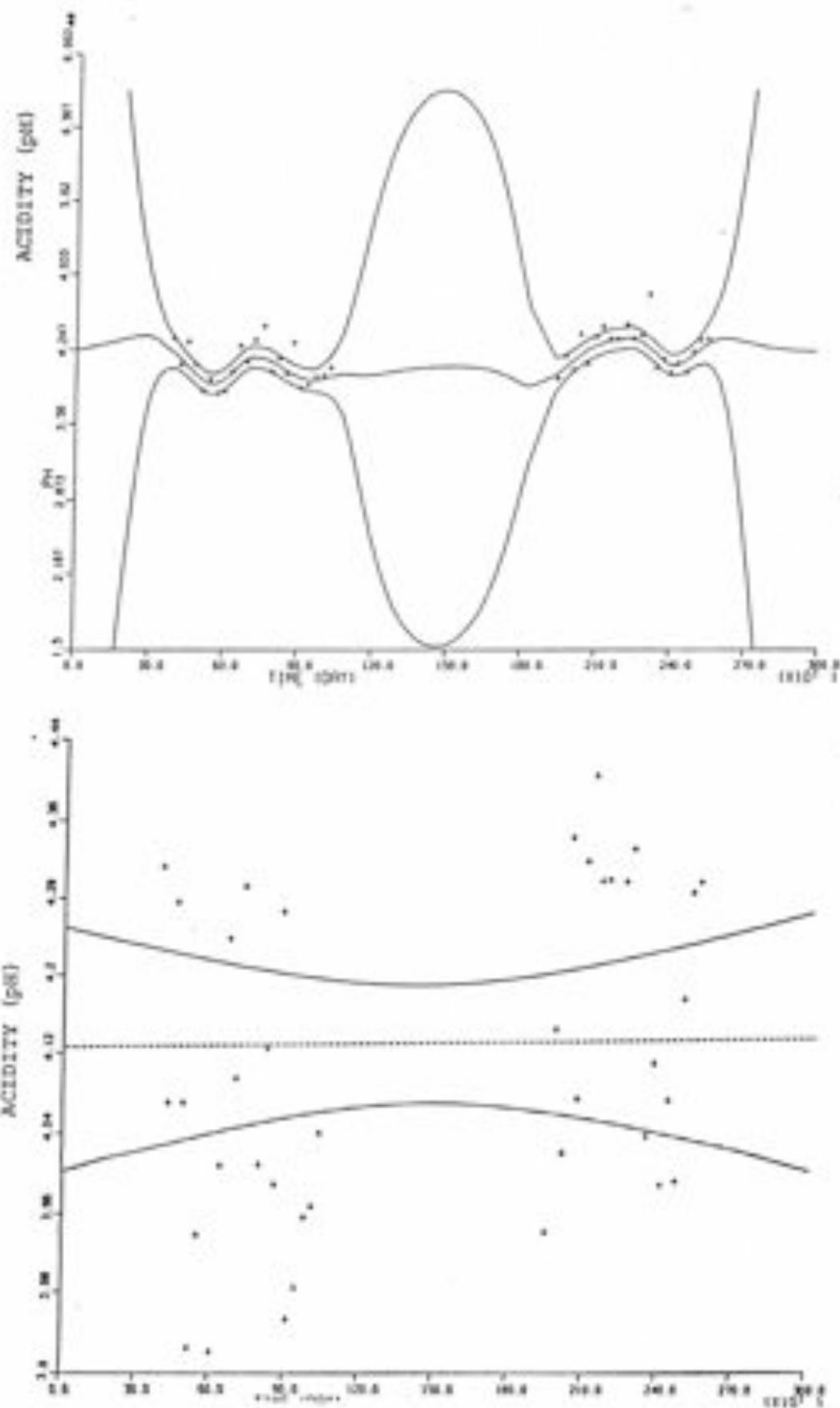


FIGURE 7. Estimated First Derivative Function for the Locally Linear Smooth and the 1977 pH Data.

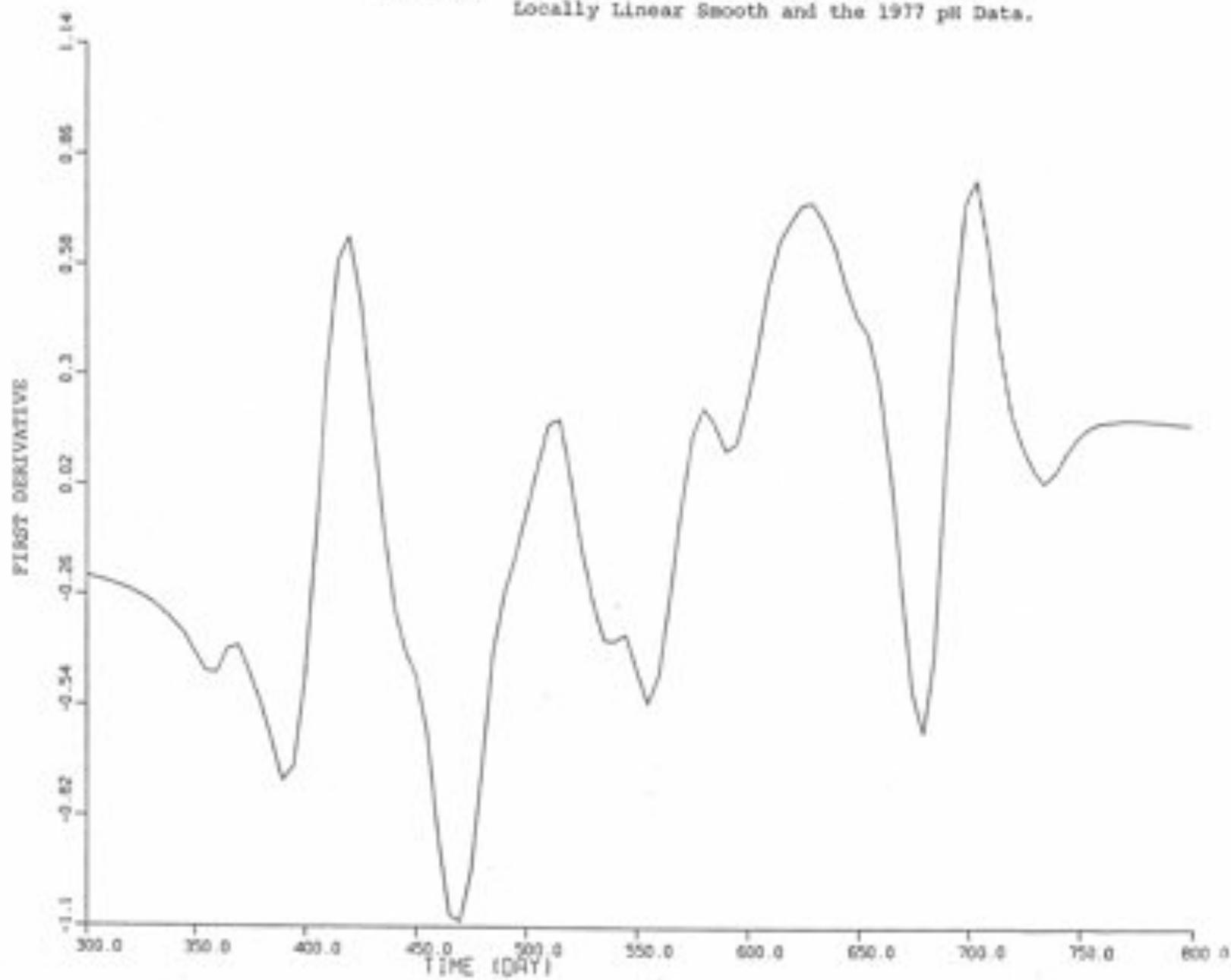


FIGURE 8. A Comparison for 1977 Data of Locally Linear Fits for Various Values of the Smoothing Constant  $c = 0.00001$ , the Smoothest (Straight line) Fit, Through  $c = 0.1, 10, 100, 1746.9$  To  $c = 10000$ , the Roughtest Fit.

