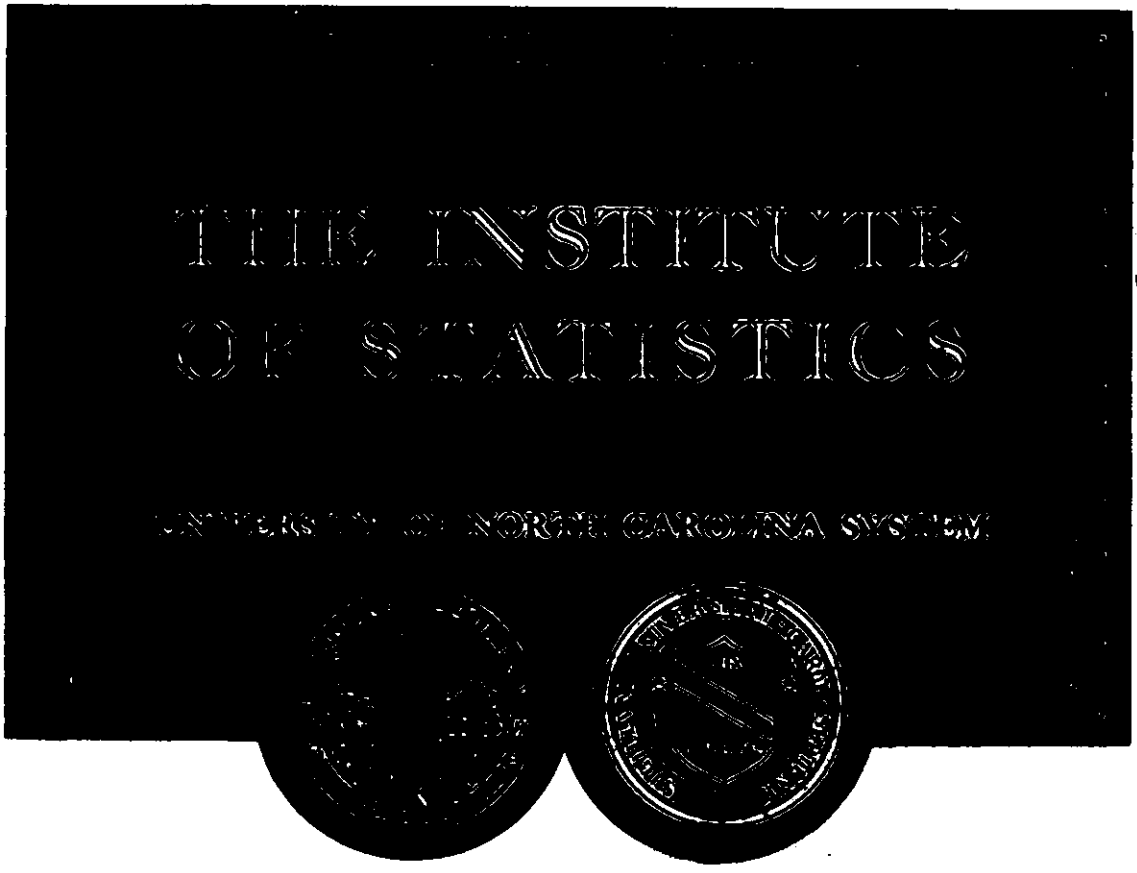


A MODEL FOR SYRINGE GRADING BASED ON EXTRACTED FEATURES FROM HIGH DIMENSIONAL FRICTION DATA

by

Michael O'Connell, Douglas Nychka,
Gerry Gray, David Martin and Perry Haaland

Institute of Statistics Mimeo Series No. 2256

May, 1993

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

M. Oconnel, D.
Nychka, G. Gray, D.
Martin & P. Haaland

A Model for Syringe Grading
Based on Extracted Features from
High Dimensional Friction Data

Name DATE

A Model for Syringe Grading Based on Extracted Features From High Dimensional Friction Data

By Michael O'Connell, Doug Nychka,
Gerry Gray, David Martin and Perry Haaland*

Abstract

In the biomedical products industry, measures of the quality of individual clinical specimens or manufacturing production units are often available in the form of high dimensional data such as continuous recordings obtained from an analytical instrument. These recordings are then examined by experts in the field who extract certain features and use these to classify individuals. To formalize and quantify this procedure an approach for extracting features from recordings based on nonparametric regression is described. These features are then used to build a classification model which incorporates the knowledge of the expert automatically.

The procedure is presented via an example involving the grading of syringes from associated friction profile data. Features of the syringe friction profiles used in the classification are extracted via smoothing splines and grades of the syringes are assigned by an expert tribologist. A nonlinear classification model is constructed to predict syringe grades based on the extracted features. The classification model makes it possible to grade syringes automatically without

*Michael O'Connell is Research Scientist II, Perry Haaland is Research Fellow and David Martin is Surface Physicist, all at Becton Dickinson Research Center, Research Triangle Park, NC 27709-2016. Doug Nychka is Professor, Department of Statistics, North Carolina State University, Raleigh NC 27695. Gerry Gray is Statistician, Fisheries Statistics Division, National Marine Fisheries Service, Silver Spring MD 20910.

expert inspection. Using leave one out cross-validation, the prediction accuracy of the classification model is found to be the same or better than the accuracy obtained from the expert.
KEY WORDS: Cross-validation; Quantile splines; Model selection.

1 INTRODUCTION

The classification of an individual production unit or clinical specimen based on a recording from an analytical instrument presents a common problem in many biomedical applications. Often, the instrument evaluating a material, sample or system, yields a continuous set of measurements which are used in the classification. Such continuous recordings must be reduced to a useful summary in order for an appropriate diagnosis to be made. Often this is done subjectively by experts in the field of application. In what follows a statistical approach to this problem is presented. The approach provides a means to automate the instrument diagnosis by building a classification model which incorporates the knowledge of the application expert.

The approach is presented by way of an application involving the quality assessment of hypodermic diabetic syringes. The syringes considered are made up of a rubber stopped piston sliding within a plastic cylinder, aided by a lubricating fluid. Syringe quality is indicated by ease of use and includes requirements of low force and smooth movement in actuating the syringe, control of subsequent syringe movement, and no leakage. Also it is important that the syringe retain its smooth operation after a pause in use.

One way of evaluating the performance quality of a syringe is to measure the amount of force necessary to produce constant acceleration of the plunger. This is achieved using an Instron^{TM} machine and results in a force/velocity curve which we refer to as a friction profile. The friction profile consists of several hundred measurements of force/velocity over a period of less than a minute. Figure 1 gives

some examples of friction profiles for three syringes of different qualities.

In general, profiles from better syringes plateau at a lower force and exhibit little oscillation after the first peak. Although the testing procedure is not a realistic simulation of syringe use, these features in the profiles are able to predict the smoothness of actuating a syringe after a pause in use. Furthermore, an engineer familiar with the testing procedure, and the hydrodynamic principles that control the frictional forces in this system plunger, can reliably convert a friction profile into a absolute grade for the syringe. In the current application syringes are graded on a soft ten point scale, with 1 being a poor syringe and 10 being the best. A tribologist can convert the friction profiles of syringes to grades with an accuracy of about $\pm$ 1 grade. The drawback to this type of procedure is that the grades are assigned subjectively based on a visual assessment. Thus the test is not generally practical for quality improvement in a manufacturing setting and also may not even be consistent among different graders.

The statistical problem is then to build a model, based on a diverse set of friction profiles, that can be used to estimate the grade of a syringe. Besides providing a rapid and objective method for grading syringes, such a model may suggest salient features in the friction profiles missed by a visual assessment. The prediction of syringe grades involves a combination of nonparametric curve fitting techniques and nonlinear regression. For each syringe the friction profile is first summarized by curve estimates describing the upper and lower boundaries. From these smooth estimates of the envelope of the friction profile, a smaller subset of features relating to the curves and their derivatives, are extracted. A classification regression model is then built to predict the visual score from the extracted features. Cross-validation is used to assess the predictive power of the model. For a diverse training sample of 51 syringes we found that the statistical classification approach is expected to be accurate to within $\pm$ 1 grade with 95% confidence.

The remainder of the article is loosely organized to follow the traditional scien-

tific format of: material and methods, results and discussion. Section 2 describes the syringe data. Section 3 describes the feature extraction process based on non-parametric regression. Section 4 describes the classification using regression models and the estimates of predictive power based on cross-validation. Some conclusions and discussion of alternative methods are given in Section 5.

2 THE FRICTION PROFILE DATA

2.1 Frictional forces in syringes

Lubricated friction can be divided into two regimes: boundary lubrication and hydrodynamic lubrication (Adamson, 1976). Boundary lubrication occurs at lower speeds when the lubricating film is penetrated by some points of the sliding surfaces, causing a relatively high frictional force. As the speed is increased the system makes a transition between boundary and hydrodynamic lubrication and this may lead to a phenomenon known as "stick-slip" or "chatter". During this period the syringe system oscillates between the high and low frictional forces associated with the boundary and hydrodynamic regimes. Hydrodynamic lubrication occurs at higher speeds when the lubricant film is not penetrated by the sliding surfaces, and the only resistance is the relatively low frictional force due to fluid shear. Thus in the hydrodynamic regime, frictional force can be adjusted by varying the lubricant viscosity.

Factors considered in the visual scoring procedure are listed in Table 1. For example, force/velocity profiles which are flat, smooth, and less than about 300 g score high. This is due to the absence of excessive boundary lubrication friction and stick-slip oscillations ($\omega = 0 \ \ \forall v$). Profiles that exhibit high forces in overcoming initial boundary friction changing to low force at high velocities ($F = f(v)$), and frequently accompanied by stick-slip oscillations ($\omega \neq 0$), are scored low. We refer to the force required to overcome initial boundary friction loosely as "break-out" force.

Three friction profiles are given in Figure 1. The associated grades cover the range from best (10) to worst (1). Profile 1 (grade=1) has a high break-out force of over 800 g and as the velocity is increased and the initial force is overcome, an oscillatory area of stick-slip behavior can be seen. This region covers a broad range of velocities, and operating a syringe in this range would result in lack of control when dispensing small amounts of medicament. At high velocity the force drops dramatically to a lower and steady level. All of the negative factors in Table 1 may be observed in profile 1.

Profile 6 (grade=6) exhibits the characteristics of profile 1 but to much less an extent. The break-out force is much lower than that observed for profile 1, but still higher than the subsequent sustaining force by at least a factor of 2. Some stick-slip oscillations are evident, though these are of lower magnitude and over a narrower range of velocities than seen in profile 1. This profile is typical of some commercially available syringes.

Profile 10 (grade=10) is an exceptionally well performing syringe. Initiation of movement requires little force, with no stick-slip throughout the range. Also the amount of force necessary to achieve constant acceleration is low and increases gradually across the range of velocities after break-out. This syringe would be easy to start and easy to control. Experience has shown that the low amplitude oscillations seen in this curve are not perceptible by a syringe user. This profile exhibits all the positive properties listed in Table 1.

2.2 Friction profile data for modeling

Initially friction profiles for three 1cc diabetic syringes were obtained and used to develop the feature extraction process. These are the profiles given in Figure 1. Subsequently, for the purpose of building the classification model, friction profiles were obtained for an additional 51 syringes. These syringes were chosen to reflect the range of quality performance anticipated from modifications to the syringe design

Table 1: Factors influencing visual scoring

positive	negative
$F < 300g$	$F > 500g$
$F \neq f(v)$	$F = f(v)$
	$F(\text{low } v) >> F(\text{high } v)$
$\omega = 0, \forall v$	$\omega \neq 0, \text{some } v$
	$\omega \neq 0, \text{wide range of } v$
F = frictional force	
v = velocity of sliding	
ω = frequency of oscillations	

in quality improvement applications. This range of syringe quality was achieved by varying the lubricant and cross-linking conditions in production of the individual syringes.

An Instron^{TM} instrument was used to obtain the friction profiles. This instrument is set up to hold the barrel of the syringe in place, and records the velocity of the plunger, and the force required to maintain constant acceleration of the plunger, as it is moved down the barrel of the syringe. The friction profile is constructed as a plot of force versus velocity.

3 ESTIMATION OF FEATURES

We now describe the statistical methods that reduce the friction profile data to a lower dimensional set of distinctive features that may be effective for grading. The nonparametric curve estimates were developed by experimenting with the friction profiles shown in Figure 1. The goal in this part of the study was to be able to reliably extract functionals of the the friction profiles that were related to the

qualitative features used to grade syringes. The functionals that were considered are all derived from three nonparametric curve estimates. These are a mean function f, and two other curves f_L and f_U that estimate a smooth envelope for the the measured force. The intent was that the mean function and f_U would be used to estimate the break out force and the average amount of force needed after the break out. The difference between the upper and lower curves was thought to be useful in quantifying the stick-slip effect. These curve estimates are described below.

3.1 Spline estimates for the mean force curve

The curve estimates considered in this work are all variations on a cubic smoothing spline where the smoothing is chosen adaptively from the data. For estimating the mean function consider the nominal model:

$$Y_k = f(v_k) + e_k \quad 1 \leq k \leq N$$

where Y_k is the force exerted when the plunger is at velocity v_k, $f(v)$ is a smooth (differentiable) function and e_k is a random component with mean zero and variance proportional to $1/w_k$. Under this model a cubic smoothing spline estimate of f is the function that minimizes:

$$(1/N)\sum_{k=1}^{N} w_k(Y_k - h(v_k))^2 + \lambda \int_{[v_1,v_N]} (h''(v))^2 dv$$

over all h such that $\int (h''(v))^2 dv$ is finite and $\lambda > 0$. The spline estimate at a particular point can be interpreted simply as weighted local average of $\{Y_k\}$ where the weights depend on the independent variable, $\{v_k\}$ and the smoothing parameter λ. Computing a cubic smoothing spline estimate is an efficient operation ($O(N)$). Also, in between observation points, the estimate has the form of a cubic polynomial and thus it is simple to evaluate the estimate or its derivatives at arbitrary points.

The choice of the smoothing parameter is important. As λ becomes very large the estimate converges to the (weighted) least squares line fit to the data. At the other

extreme as $\lambda \to 0$, the estimate is a function that interpolates the data. Intermediate choices for λ strike a balance between fitting the data well and constraining the resulting estimate to be a smooth curve. Often the smoothing parameter is chosen subjectively by examining the data. However, since the interest in this study is in providing an objective procedure for grading, a data-based method is used to estimate λ. For fixed λ and $\{v_k\}$, the spline estimate is a linear function of $\{Y_k\}$. If $\hat{f}_\lambda$ denotes the spline estimate then there is a "hat" matrix $A(\lambda)$ such that ($\hat{f}_\lambda(v_1)$, ..., $\hat{f}_\lambda(v_N))^T = A(\lambda)Y$. With this notation the generalized cross-validation (GCV) function is

$$V(\lambda) = \frac{(1/N)[RSS(\lambda)]}{[1 - m(\lambda)/N]^2}$$

where $RSS(\lambda)$ is the residual sum of squares as a function of the smoothing parameter and $m(\lambda) = trA(\lambda)$ is one measure for the effective number of parameters needed to describe the curve estimate. A suitable value for the smoothing parameter is taken as the value that minimizes $V(\lambda)$ over all λ in the range $[0, \infty]$. For some background on splines and other smoothing methods see Eubank (1988) and Hastie and Tibshirani (1990, Chapter 3).

3.2 Smooth estimates of the upper and lower boundaries

Two methods for estimating the functional form of the lower and upper boundaries, f_L and f_U, are developed. The first method, which we refer to as extremal splines, is based on estimating the upper boundary by smoothing subsets of the data that are local maxima. The lower curve is found in a similar manner by considering local minimum. The second strategy estimates the 5% and 95% conditional quantiles of the friction profile data and is referred to as quantile splines.

To estimate an envelope that contains the limits of the force applied to the syringe one can take advantage of the physical properties of these data. Although the response for a poor syringe might be represented as an additive model, possibly with a changing variance function, it is more appropriate to think of these data

as the result of under-sampling a rapidly oscillating function. In these tests the measurement error is actually negligible and the variation in the amount of force is due to interactions among the rubber in the plunger, the silicon lubricant and the barrel of the syringe. For this reason friction profiles such as those in Figure 1 can be interpreted as points from a continuous function where the sampling rate is on the order of the higher frequency oscillations. Were it possible to measure the force in smaller increments of the velocity, the trajectory would tend to trace out a smooth but rapidly varying curve. From this perspective the envelope of the friction profile should be related to the amplitude of these oscillations. Thus one method for calculating such an envelope would be to find points that are local maxima and minima, and then fit smooth curves through these points. A simple definition of a local maximum (minimum) is a point which is larger (smaller) than each of its two nearest neighbors. That is, based on the original velocity/force pairs, form the subsets:

$$\{v_{j,U}, Y_{j,U}\} : \text{the set of } \{v_k, Y_k\} \text{where } Y_{k-1}, Y_{k+1} < Y_k$$

$$\{v_{j,L}, Y_{j,L}\} : \text{the set of } \{v_k, Y_k\} \text{where } Y_{k-1}, Y_{k+1} > Y_k$$

Upper and lower envelope curves are now estimated by fitting smoothing splines to these sets of local minima and maxima. One advantage of this approach is that the usual cross-validation function can be used to select smoothing parameters.

Figure 2 gives the estimated extremal spline for the friction profile of the grade 1 syringe. It is our experience that the extremal spline estimates were found to work well in capturing one's visual impression of the features that affect a syringe's grade. One problem with these estimates, however, is the bias in describing the sharp rise in the friction profiles before breakout. This problem is due to the fact that no local minima or maxima occur in this range of the data and thus the estimated envelope is based on extrapolation from larger velocities.

As an alternative, quantile splines (Bloomfield and Steiger, 1983), were also considered. These splines allow estimation of extreme quantiles of the data and

thus capture some form of smooth boundary. Let

$$\rho_\alpha(u) = \begin{cases} \alpha\,|u| & \text{if } u \le 0 \\ (1-\alpha)\,|u| & \text{if } u > 0 \end{cases}$$

For any vector Z it well known that the α^{th} quantile is a minimizer of $\sum_{k=1}^{n} \rho_\alpha(Z_k - q)$ over all values of q. Based on this characterization a quantile spline is taken as the minimizer of

$$(1/N)\sum_{k=1}^{N} \rho_\alpha(Y_k - h(v_k)) + \lambda \int_{[v_1, v_N]} (h''(v))^2 dv$$

over all h such that the roughness penalty is finite. Although a quantile spline is a nonlinear function of the data, an approximate minimizer can be calculated by a iterative procedure based on weighted least squares smoothing splines. Since the least squares spline can be calculated rapidly, the computational burden for determining a quantile spline is not large. (See the Appendix for details.)

An appropriate choice of smoothing parameter may vary as a function of α and for this reason it is important to estimate λ adaptively from the data. One strategy for choosing λ is an approximate form of leave-one-out cross-validation. Let $\hat{f}^{[k]}_{\lambda,\alpha}$ denote the quantile spline having omitted the k^{th} data point, (v_k, Y_k). Now suppose that $\hat{f}^{[k]}_{\lambda,\alpha}$ has been computed for each k then one might consider the quantile cross-validation criterion:

$$QCV(\lambda) = (1/N)\sum_{k=1}^{N} \rho_\alpha(Y_k - \hat{f}^{[k]}_{\lambda,\alpha}(v_k))$$

as a measure of the fit of the quantile spline to the data. For a particular α, an estimate of λ is obtained as the smoothing parameter that minimizes QCV. Because the quantile spline is a nonlinear function of the data it is computationally intensive to determine QCV exactly. However, one can approximate $\hat{f}^{[k]}_{\lambda,\alpha}$ with the result of just one iteration of the weighted least squares algorithm using the estimate for the full data as the initial value (see Appendix). This strategy may not be accurate if a single data point has a large influence on the quantile estimate. However, for

the sample sizes encountered in this application we found that approximate cross-validation was stable and gave reasonable results.

Figure 2 also shows the 5% and 95% quantile spline estimates for the friction profile data where λ has been found by the approximate cross-validation procedure mentioned above. Note that in contrast to the envelope splines these quantile estimates give a sharper characterization of the boundary in the initial part of the curve. One reason for this difference is that the quantile estimates use the full data while the extremal splines are restricted to a subset (either local maxima or minima). Because the first part of the friction profile is monotonic, no points in this range are used in deriving the extremal splines.

It was also found that the quantile estimates based on cross-validation tend to smooth the data more than the extremal splines. This property resulted in more stable estimates of the derivatives of the force/velocity curves. Section 4.1 and Figure 6 discuss this issue further.

3.3 Extracting distinctive features from the friction profiles

Based on the visual features used to grade a syringe, we considered the following functionals based on the mean and boundary curves. They are numbered in accordance with later figures regarding model selection.

1 Effective number of parameters for $\hat{f}$, where the smoothing parameter is chosen by cross-validation.

2 Maximum difference between $\hat{f}_U$ and $\hat{f}_L$.

3 Mean difference between $\hat{f}_U$ and $\hat{f}_L$.

4 Velocity at minimum second derivative of $\hat{f}_U$.

5 Minimum second derivative of $\hat{f}_U$.

6 Force at minimum second derivative of $\hat{f}_U$.

7 Velocity at first maximum of $\hat{f}_U$ (breakout point), V_{BP} .

8 Force at first maximum of $\hat{f}_U$, F_{BP} .

9 25th percentile of $\{ \hat{f}'(v_k)\}$ for $v_k \geq V_{BP}$.

A 50th percentile of $\{ \hat{f}'(v_k)\}$ for $v_k \geq V_{BP}$.

B 75th percentile of $\{ \hat{f}'(v_k)\}$ for $v_k \geq V_{BP}$.

The first functional is an overall measure of the complexity of the estimated mean function. The next two functionals are measures of the difference between the upper and lower curves and are intended to quantify the stick-slip effect. There are two strategies to identify the breakout point. One is to find the point of maximum curvature in the upper envelope (4-6) and the other is to identify the first local maximum in the test (7 and 8). The last three measures attempt to summarize the smoothness of the profile after the breakpoint and the trend of the profile. Recall that a syringe with a high grade tends to have a smooth but slightly increasing force trajectory after the breakout point. Smoothness in this region may be characterized by the scatter in the first derivative of the mean curve while a gradual trend in the profile is indicated by a positive median first derivative.

4 CLASSIFICATION OF SYRINGES FROM A TRAINING SAMPLE

4.1 Selecting a subset of features

For each of the 51 syringes tested mean, upper and lower boundary curves were estimated using extremal and quantile splines. Based on these curves the 11 summary functionals were computed. The first step was to identify a subset of the functionals that gave good linear predictions of the actual grades. We first consider

classification models of the form

$$Z_k = X^T_{[k,]}\beta + e_k \tag{1}$$

where Z_k is the grade of the kth syringe as assigned by the tribologist; $X^T_{[k,]}$ is the kth row of X^T which corresponds to the extracted features of the kth syringe; β is the vector of coefficients corresponding to the features and e_k is a random error term. Since the main interest is in the predictive power of these summary measures, Mallow's C_p statistic was used to identify a subset of the functionals. Figure 3 includes a plot of the C_p statistics for the different subsets of features based on extremal splines. This model selection criterion indicates that a linear model based on 8 features is the best choice for the extremal splines approach. It is interesting to note that the best three variable model includes the basic visual aspects of grading: the stick-slip effect noticed for poor syringes (3), the break out force as indicated by the force at the minimum second derivative of $\hat{f}_U$ (6), and the general trend in the friction profile after breakout (A).

Besides using Mallow's criterion, generalized cross validation (GCV) was also considered. For linear regression the GCV function for a subset of size p is $V(p) = (1/n)RSS(p)/(1-p/n)^2$. Under the assumption that p/n is small and $RSS(p)/(n-p)$ is approximately equal to $\hat{\sigma}^2$ based on the full set of variables, $V(p)$ is similar to $\hat{\sigma}^2 C_P$. However when p/n is large GCV is more conservative and penalizes larger models more heavily. For this reason the GCV criterion was also considered for subset selection. Figure 3 also includes a plot of the GCV statistics for the different subsets of features based on extremal splines.

For features derived from the extremal splines the GCV and Cp criteria (Figure 3) indicate a local minima ($GCV \doteq .5$) at 5 features. The additional two features are: the effective number of parameters for $\hat{f}$ (1), and the minimum second derivative of $\hat{f}_U$ itself (5). In the interest of parsimony this was the subset of features considered in subsequent analysis. It should be noted that the GCV criterion is a nearly unbiased estimate of the expected average square prediction error. The best five feature

subset to grade syringes has a prediction standard deviation of approximately 0.7. Thus the linear model using these five features can identify the syringe grade to within ± 1.4 with about 95% confidence.

The best subset of features based on quantile splines has four members namely 3, 8, A and B as indicated in Figure 4. This model has a slightly higher prediction error compared to the extremal spline model discussed above.

The two linear classification models showed a substantial difference in predicting the grades of two grade 3 syringes (Figure 5). The grade for these two cases was predicted well by the extremal features and not well by the quantile features. Figure 6 shows the friction profile for one of these syringes and the estimated envelopes for the two methods. The results are counter to what one would expect in that although the quantile spline doesn't estimate the grade as well, it does appear to estimate the features more accurately than the extremal spline. The extremal spline undersmooths the set of local maxima to give a poor estimate of the upper boundary. The rough upper boundary estimate translates into an unstable estimate of the breakout force based on the second derivative. This breakout force estimate of $F_{BP} \doteq \hat{f}(v = 0.05) \doteq 1.25$, is substantially lower than what one would estimate from just a visual inspection of the force/velocity curve ($F_{BP} \doteq 1.6$). In contrast the quantile spline yields a smooth estimate of the upper boundary and appears to give an accurate estimate of the break out force for this syringe. Given the discrepancy between these two estimates of the envelope it is not surprising that the predicted grades differ. However, it is unusual that the poorer extremal curve estimate does better. One simple explanation is that the syringe was incorrectly graded by the expert and more will be said about this in the discussion.

4.2 Nonlinear modifications of the model

For all of the linear models based on subset selection, the residuals indicated some systematic departures from a linear function. The residual plots for the best quan-

tile spline model (Figure 7) and the best five feature subset extremal spline model (Figure 8) suggest that agreement with the actual grades can be improved if the estimated grades are transformed. Because the grades are qualitative there is no reason to assume the unit differences among grades correspond to equal spacing in the curve features. We therefore consider the nonlinear model

$$Z_k = H(X_k^T \beta) + e_k \tag{2}$$

where Z_k is the grade and $(X_{k1}, ..., X_{kp})$ are the functionals estimated from the k^{th} friction profile and H is a smooth transformation. A direct way for estimating H is to fit a smooth curve to the pairs $(\hat{Z}_k, Z_k)$ where $\hat{Z}_k$ are the predicted values based on fitting a linear model. Figures 7 and 8 summarize the results of estimating H in this way using a smoothing spline. Note one adjustment made by the transformation for both the quantile and extremal spline models is to discount predicted grades in the range of 4 to 6. The residuals for the transformed models (Figure 9) appear to have less dependence on the predicted values than the untransformed linear model residuals (Figures 7 and 8) but there is still some bias in estimating very high grades. After estimating the transformation the parameter vector β was re-estimated using nonlinear least squares, with the transformation considered fixed. From a practical point of view this refinement of the parameter estimates gave little improvement.

Several other models were also considered. Given the large range for some of the features, such as the average difference in the upper and lower envelopes, it is unlikely that a linear model for the features is optimal even when the results are sharpened by a nonlinear transformation. In order to explore some alternatives we tried an additive model based on the best subsets indicated by the GCV criterion. Additive models of the form:

$$Z_k = \sum_{j=1}^{p} f_j(X_{k,j}) + e_k$$

were found to have the same bias in the residuals observed in the linear models (see Figure 7). Another variation on the transformation model is to transform the

dependent variable rather than the prediction equation:

$$G(Z_k) = X_k^T \beta + e_k$$

This model can be estimated using the same alternating strategy for determining an ACE model (Friedman and Breiman 1985). Surprisingly this model did not adequately adjust for the nonlinear effect in predicting the intermediate grades and gave a similar residual pattern to the simple linear model.

An extension of this ACE-type model is

$$h(Z_k) = \sum_{j=1}^{p} f_j(X_{k,j}) + e_k.$$

This model combines both the nonparametric flexibility of how features enter the model along with the chance to consider transformations of the grades. We did not fit this model due to the small size of the training sample.

4.3 Estimating prediction error

The final aspect of our analysis of the training sample was to assess the prediction error from the nonlinear model that incorporates a transformation (model (1)). Since the number of data points is small (51) the prediction error was estimated by cross-validation. Each syringe was sequentially omitted from the data set and model (??) was fit to the remaining data. The estimated model was then applied to the omitted point to predict the syringe grade. Let $\hat{Z}_k^{[k]}$ denote this prediction. Note that Z_k is independent of $\hat{Z}_k^{[k]}$ and the mean square of the cross-validation residuals is an estimate of the mean square prediction error. Also the CV residuals themselves are useful for model checking. For this training sample of 51 syringes, the mean prediction error was estimated to be $(0.638)^2$ based on extremal spline features and $(0.642)^2$ based on quantile splines on the transformed scale. Not surprisingly these results are an improvement over assigning grades based on linear models. Figure 10 is a plot of the cross-validated residuals and is further confirmation that

the transformation of the linear grade reduces the bias for predicting intermediate grades (compare with Figures 7 and 8). One systematic pattern in these plots is that the model does not do an adequate job in distinguishing between syringes of grades 9 and 10.

5 DISCUSSION

The results from modeling the training sample of friction profiles indicate that syringes can be classified with a average accuracy of approximately ± 1 grade with 95% confidence. Since visual evaluation by a trained engineer is only expected to be this accurate we believe that this prediction makes a significant contribution to the automatic grading of syringes.

Our strategy for predicting syringe grades combined four basic methods: non-parametric regression techniques to extract distinctive features of the friction profiles, variable selection for identifying a parsimonious subset of features, nonlinear regression to estimate a predictive classification model and finally leave-one-out cross-validation of this model to estimate the average prediction error. Although some of the details in this are necessarily specific to the friction profile measurements we believe that this work may serve as a blueprint for other studies that involve predicting semi-qualitative properties from continuous recordings produced by an instrument. Perhaps the most difficult part in this process is identifying potential features of the instrument recording that may be useful for prediction. In the syringe application these features could be motivated in part by the physical system.

Despite a low average prediction error there are several minor problems with the final prediction models. Based on the features extracted from the friction profiles it does not seem possible to distinguish among syringes with grades 9 or 10. We choose not to refine the model because from a practical point of view the differences among these very good syringes is not important. Another problem is the large

residuals observed for the two syringes of grade 3 when the features are based on quantile splines. Besides affecting the prediction error these two also influence the smoothness of transformation estimate, H. Examining Figure 7 one may note the $\hat{H}$ dips slightly in the range $[2, 4]$ in an attempt to compensate for these two points. We decided to leave these data in the model so that estimates of the prediction error would be conservative.

REFERENCES

Adamson, A. W. (1976), *Physical Chemistry of Surfaces*, New York: John Wiley.

Bloomfield, P., and Steiger, W. (1983), *Least absolute deviations: theory, applications and algorithms*, Boston: Birkhauser.

Brieman, L., and Friedman, J. H. (1985), "Estimating optimal transformations for multiple regression and correlation" (with comments), *Journal of the American Statistical Association* 80, 580-619.

Eubank, R. (1988), *Smoothing Splines and Nonparametric Regression.* New York: Marcel Dekker.

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models.* New York: Chapman and Hall.

Reinsch, C. (1967), "Smoothing by spline functions". *Numer. Math.* 10, 177-83.

APPENDIX: QUANTILE SPLINES

This appendix describes the iterative technique for computing a quantile spline and also defines the approximate cross-validation function associated with this estimate.

5.1 Background for the Computational Algorithm

It is first necessary to present an alternative characterization of a cubic smoothing spline. Throughout this discussion we will assume that at least three of the observation points, v_k , are unique. From basic theory for splines one can show that the function that minimizes the spline objective function must be a piecewise cubic polynomial with join points (knots) at v_k , $1 \leq k \leq N$ and 'natural' boundary conditions. Based on this form, it is enough to know the value of the spline at the knot points. The rest of the values in between knots can be inferred from the piecewise cubic form. Thus, an alternative version of the usual spline minimization problem is

$$min_{h \in \Re^n}(1/N)\sum_{k=1}^{N}(Y_k - h_k)^2 w_k + \lambda h^t R h \qquad \text{(A)}$$

where $h_k = h(t_k)$ and R is an $N \times N$ matrix that can be derived from the roughness penalty and the cubic spline basis. (For this discussion, it is not necessary to give an explicit formula for R.)

The solution to (A) is equivalent to solving the linear system:

$$2(Y_j - h_j)w_j + \lambda(Rh)_j = 0 \qquad \text{(B)}$$

for $1 \leq j \leq N$. This solution can be found very efficiently by taking advantage of the banded structure of the matrix R. The best known algorithm is that of Reinsch (1967).

The strategy for computing a (approximate) quantile spline estimate is based on an iterative procedure where each step involves the solution to a system of equations like those in (B). In order to carry this out, however, it is necessary to make a slight modification to ρ_α so that it is differentiable at zero. The intuitive idea is round out the corner of ρ_α at zero by piecing in a quadratic function. Accordingly let

$$\rho_\alpha(u) = \begin{cases} \rho_\alpha & \text{for } \mid u \mid > \epsilon \\ \alpha u^2/\epsilon & \text{for } 0 \leq u \leq \epsilon \\ (1-\alpha)u^2/\epsilon & \text{for } -\epsilon \leq u \leq 0 \end{cases}$$

When $\alpha = .5$, $\rho_{\alpha,\epsilon}$ is the Huber weight function used in robust regression. The important difference is that in our calculations ϵ is chosen to be effectively zero relative to the magnitude of the data values. In robust regression, ϵ is usually of the same order as the residuals.

If this approximation is applied to the minimization functional below, one can characterize the solution in a similar manner to the weighted least squares spline in (A). Consider

$$(1/N)\sum_{k=1}^{N}\rho_{\alpha,\epsilon}(Y_k - h(v_k)) + \lambda \int_{[v_1,v_N]} (h''(v))^2 dv \qquad (3)$$

Representing the solution in terms of the value of the function at the observation points, a necessary condition for achieving the minimum value of (3) is

$$\psi_{\alpha,\epsilon}(Y_j - h_j) + \lambda(Rh)_j = 0 \quad 1 \le j \le n \qquad \text{(C)}$$

where $\psi_{\alpha,\epsilon} = \rho'_{\alpha,\epsilon}$. Moreover, with at least three unique knots it is known that a unique minimum of (3) will always exist. This result implies that (C) provides a sufficient set of conditions for the solution. Note that (C) is a nonlinear set of equations but has a similar form to those associated with the weighted least squares spline.

Let h^0 denote an initial starting value for the solution to (C). Rewrite the system as

$$2(Y_k - h_k)\left[\frac{\psi_{\alpha,\epsilon}(Y_k - h_k)}{2(Y_k - h_k)}\right] + \lambda(Rh)_k = 0 \qquad \text{(D)}$$

and now substitute the intial estimate for the occurrences of h_k in the bracketed term. With this substitution one obtains a linear system of equations in h. Identifying w_k with $[\psi_{\alpha,\epsilon}(Y_k - h_k^0)/2(Y_k - h_k^0)]$ one can solve this approximate system using the standard algorithm for weighted least squares smoothing splines. In general if h^J is the vector at the J^{th} iteration then h^{J+1} is obtained by solving the linear

system where the weights are based on h^J. Note by construction of the weighting function, at convergence h^∞ will solve (C).

In our computations the quantile splines were estimated for a grid of smoothing parameters and the estimates were computed in descending order with respect to λ. An ordinary least squares spline was used as the starting value for computing the estimate with the largest value of the smoothing parameter. Computations at other values of λ use the previous quantile estimate as a starting value. Since the quantile spline is a continuous function of the smoothing parameter, using the previous estimate can significantly reduce the number of iterations needed for convergence.

5.2 Approximate Cross-validation

The approximate cross-validation for quantile splines is based on the approximation of the estimate $\hat{h}$ by a particular weighted least squares smoothing spline. For weighted least squares splines there is a simple formula for leave-one-out cross-validation of the estimate. Let $\hat{f}_\lambda^{[k]}$ denote the least squares spline estimate in (A) but having omitted the k^{th} data point, Y_k. It can be shown (Craven and Wahba 1979) that

$$Y_k - \hat{f}_\lambda^{[k]}(v_k) = \frac{Y_k - \hat{f}_{\lambda^*}(v_k)}{1 - A_{kk}(\lambda^*)}$$

where A is the "hat" matrix discussed in Section 3.1 and $\hat{f}_{\lambda^*}$ is the spline estimate based on the full data with smoothing parameter $\lambda^* = \frac{n-1}{n}\lambda$.

Unfortunately since the quantile spline estimate is a nonlinear function of Y, it does not reduce to such a simple form. However, recall that at convergence of the iterative algorithm the quantile spline is the solution to an associated weighted least squares spline problem. If $A_w(\lambda)$ is the associated 'hat' matrix for the quantile estimate at convergence, then one could consider the approximate leave-one-out estimate:

$$Y_k - \hat{f}_\lambda^{[k]}(v_k) \simeq \frac{Y_k - \hat{f}_{\lambda^*}(v_k)}{1 - A_{w,kk}(\lambda^*)}$$

This form has an attractive interpretation in terms of the computational algorithm. Using the quantile spline for the full data as the starting vector, the approximation is the result of doing just one iteration of the algorithm on the reduced data set. This is a reasonable approximation if the quantile estimate for the full data set is not particularly sensitive to the omitted data point. With good starting values it is our experience that the convergence of the iteratively reweighted least squares procedure is rapid. Thus a single iteration should be adequate in terms of approximating the leave-one-out estimate for the purpose of constructing the comparative statistic

$$QCV(\lambda) = (1/N)\sum_{k=1}^{N} \rho_\alpha(Y_k - \hat{f}^{[k]}_{\lambda,\alpha}(v_k)) \, .$$

FIGURE LEGENDS

Figure 1. Examples of friction profiles. Friction profiles for grade 1, grade 6 and grade 10 syringes. Force is force required to maintain constant acceleration and is plotted against measured velocity.

Figure 2. Extremal and quantile spline fits to a grade 1 friction profile, (a) extremal splines, (b) 5% and 95% quantile splines. Upper and lower envelopes are estimated using extremal and quantile splines.

Figure 3. Linear classification model selection: features from extremal splines, (a) model selection based on GCV and (b) on Cp statistics. For each number of features in the model the combination with the lowest value of the model selection statistics is labeled with the appropriate feature labels (see text). Labels are centered at the number of features. Other combinations are denoted with a + symbol.

Figure 4. Linear classification model selection: features from quantile splines, (a) model selection based on GCV and (b) on Cp statistics. For each number of features in the model the combination with the lowest value of the model selection statistics is labeled with the feature numbers (see text). Labels are centered at the number of features. Other combinations are denoted with a + symbol.

Figure 5. Residual plots for linear models, (a) residuals from both models, (b) residual difference. Residuals from linear models based on extremal spline features are plotted against those from linear models based on quantile spline features. Note the two grade 3 syringes that were predicted differently by the two models.

Figure 6. Comparison of extremal and quantile spline estimates of a grade 3 syringe, (a) estimated envelopes based on extremal and (b) on quantile splines, (c) estimated

second derivatives for the upper boundary estimates. Break-out forces based on the two estimated second derivatives are marked. Note the smooth estimated second derivative based on the quantile spline and the resulting accurate estimate of the break-out force.

Figure 7. Non-linear classification model construction: features from extremal splines, (a) residual plot for the linear model, (b) spline estimate of the transformation. The spline estimate of the transformation is obtained from a spline fit to the pairs $(\hat{Z}_k, Z_k)$, where $\hat{Z}_k$ are the predicted grades based on the 5 term linear model of features obtained from extremal spline fits to the friction profiles.

Figure 8. Non-linear classification model construction: features from quantile splines, (a) residual plot for the linear model, (b) spline estimate of the transformation. The spline estimate of the transformation is obtained from a spline fit to the pairs $(\hat{Z}_k, Z_k)$, where $\hat{Z}_k$ are the predicted grades based on the 4 term linear model of features obtained from quantile spline fits to the friction profiles.

Figure 9. Residual plots for non-linear models, (a) extremal spline model, (b) quantile spline model. Note the better overall fit, and in particular the better fit between grades 4 and 8, of these spline transformed models as compared to the residual plots given in Figures 7 and 8.

Figure 10. Cross-validated residual plots for non-linear models, (a) extremal spline model, (b) quantile spline model. Note the similar patterns to those in Figure 9.

Figure 1. Examples of friction profiles

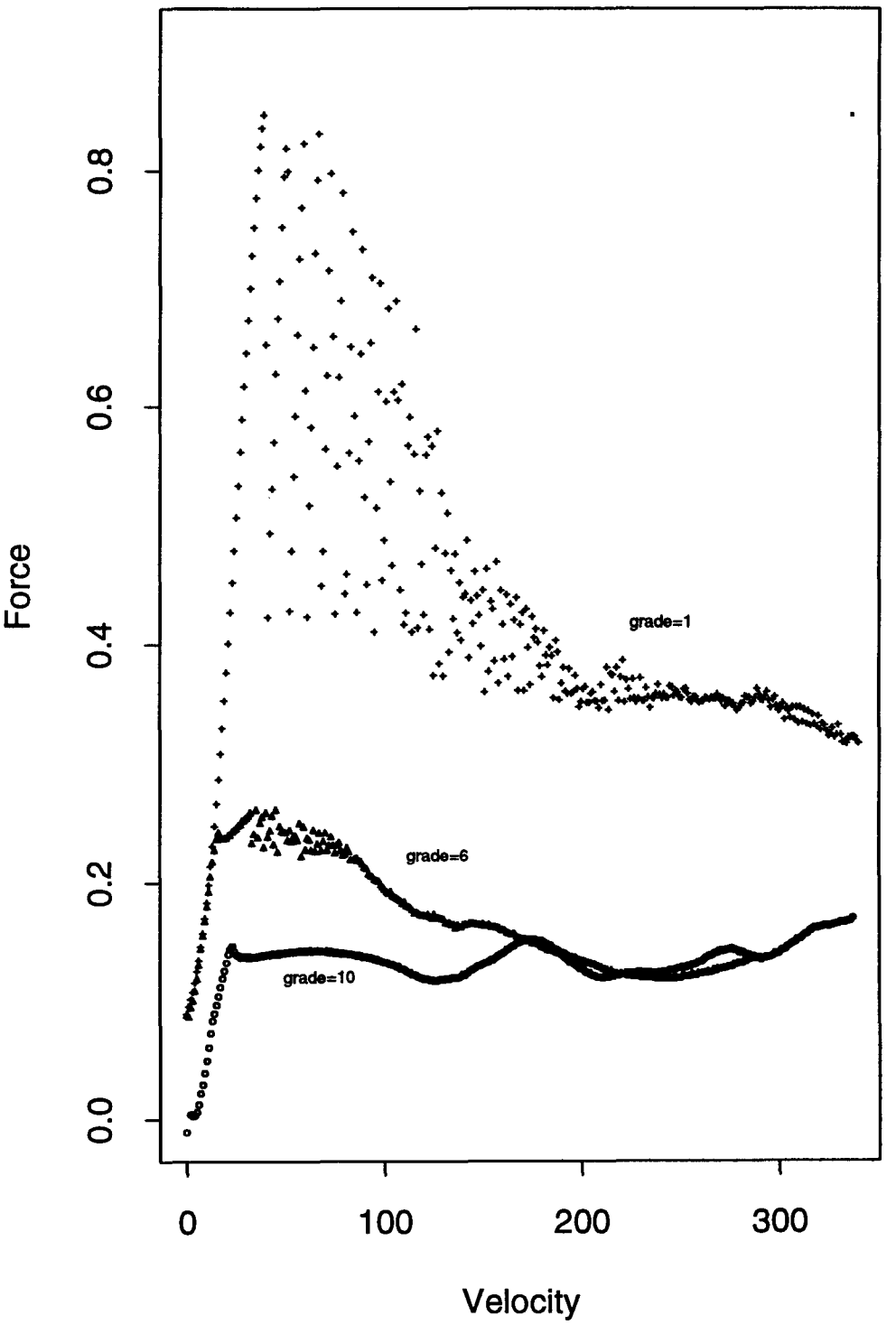
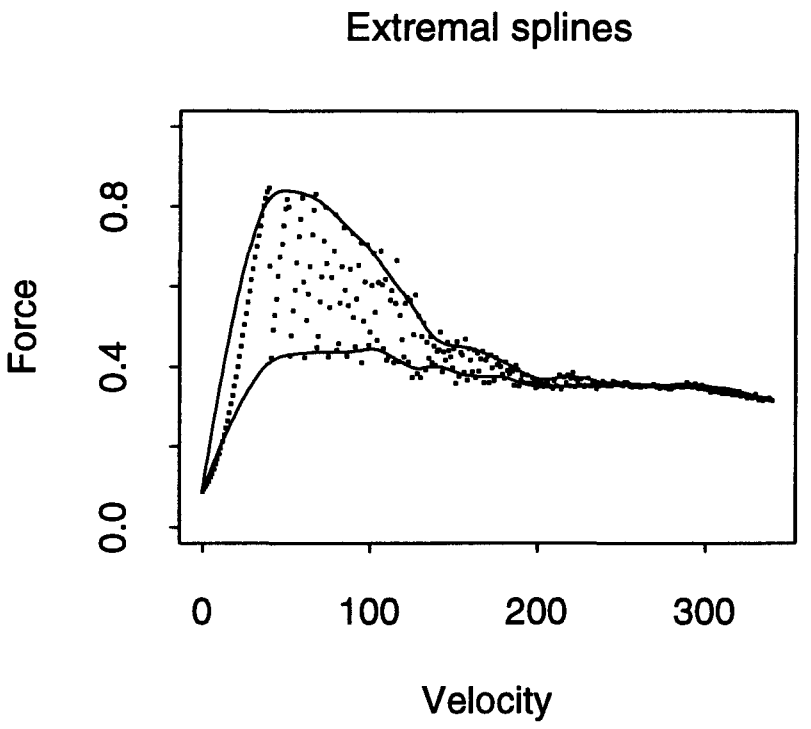


Figure 2. Extremal and quantile spline fits to a grade 1 friction profile



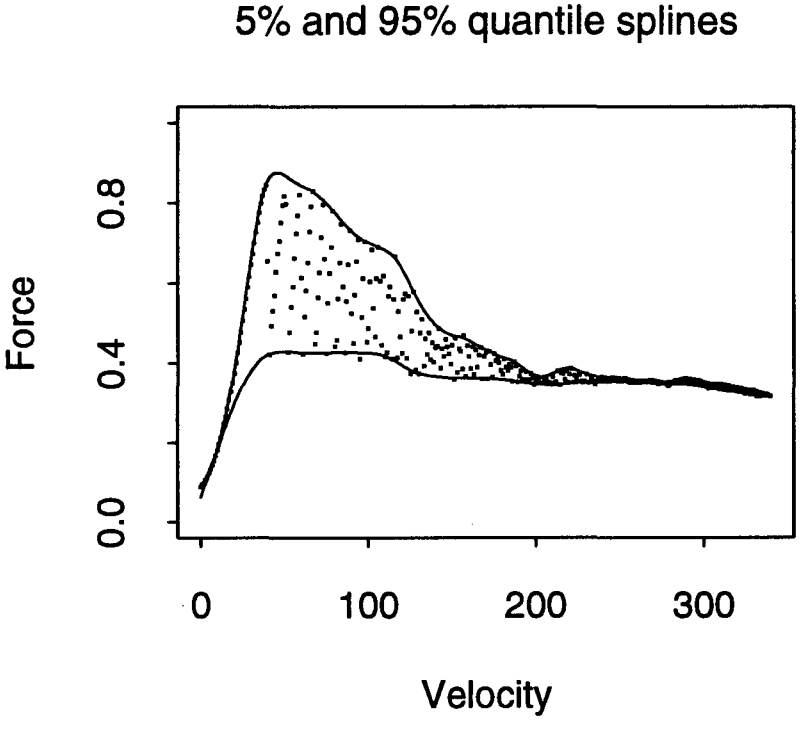
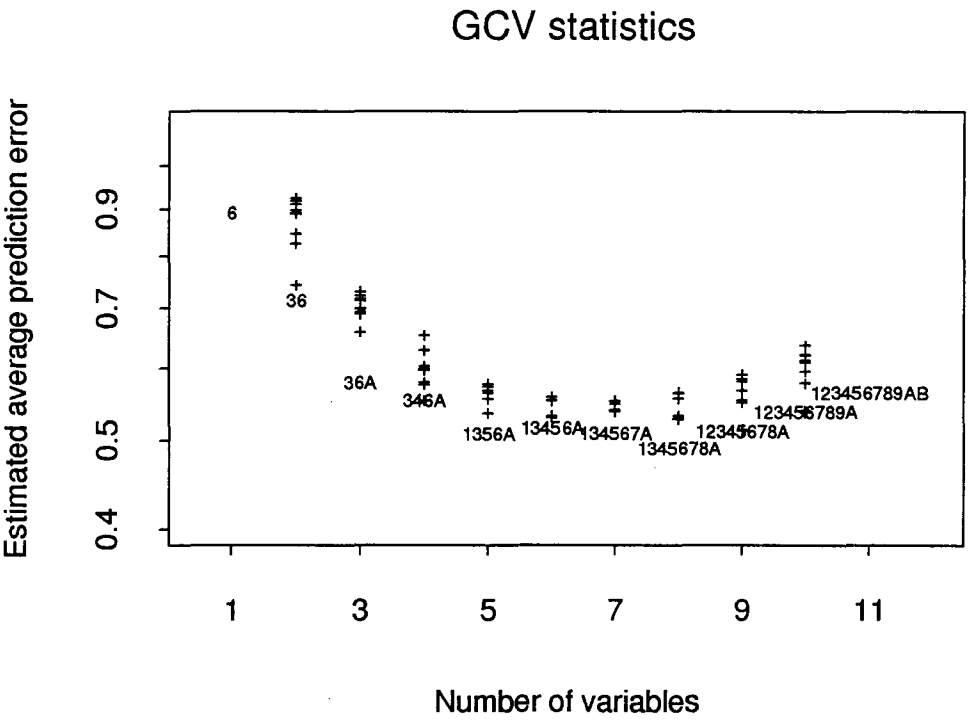


Figure 3. Linear classification model selection: features from extremal splines



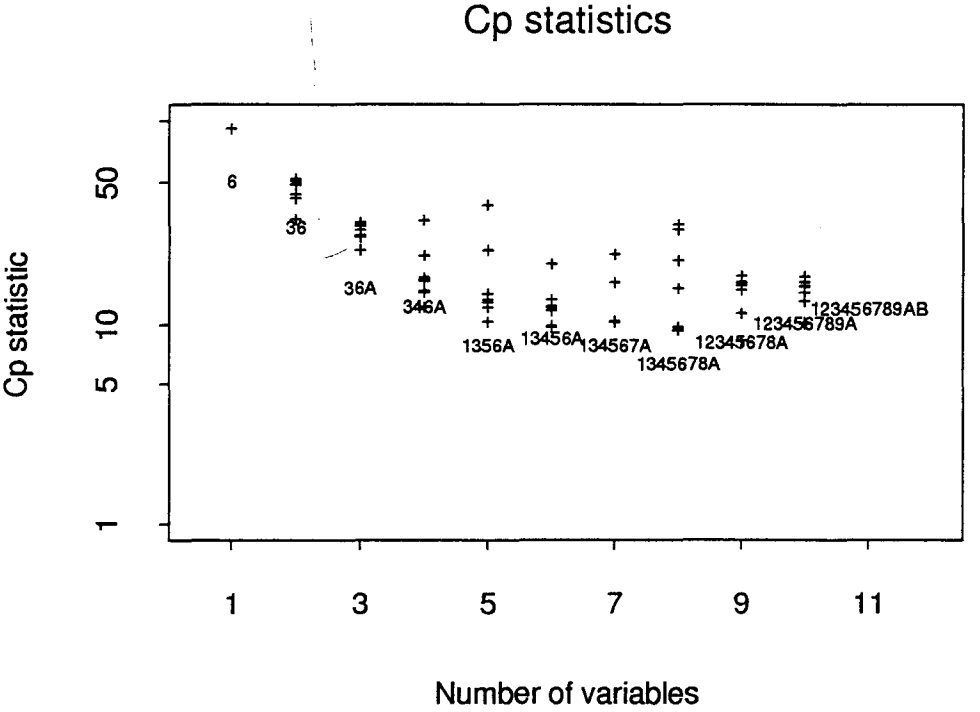
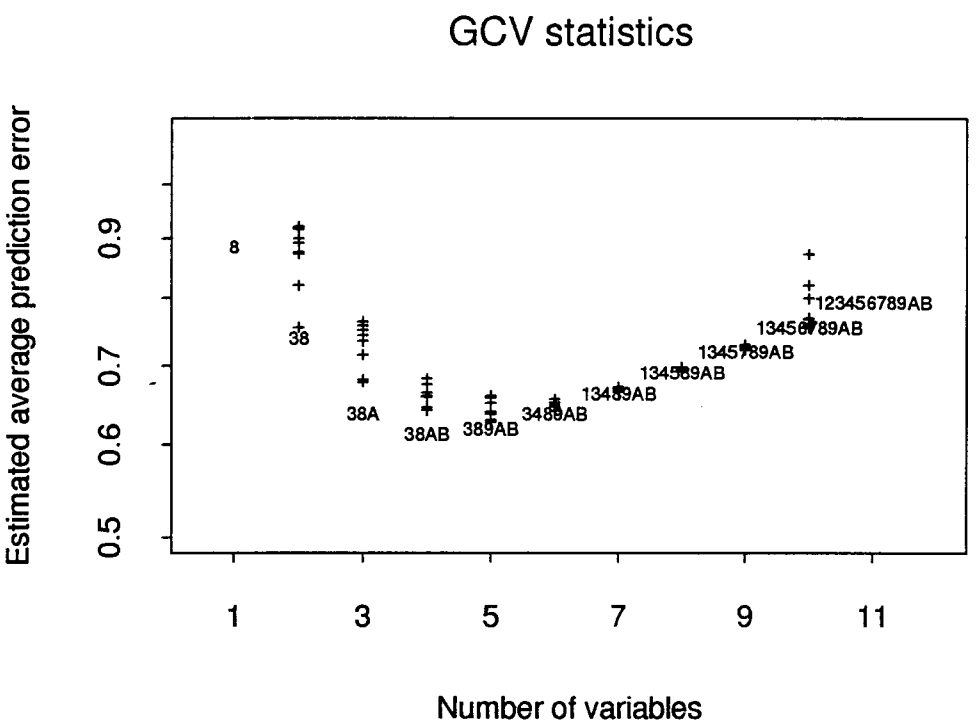


Figure 4. Linear classification model selection: features from quantile splines



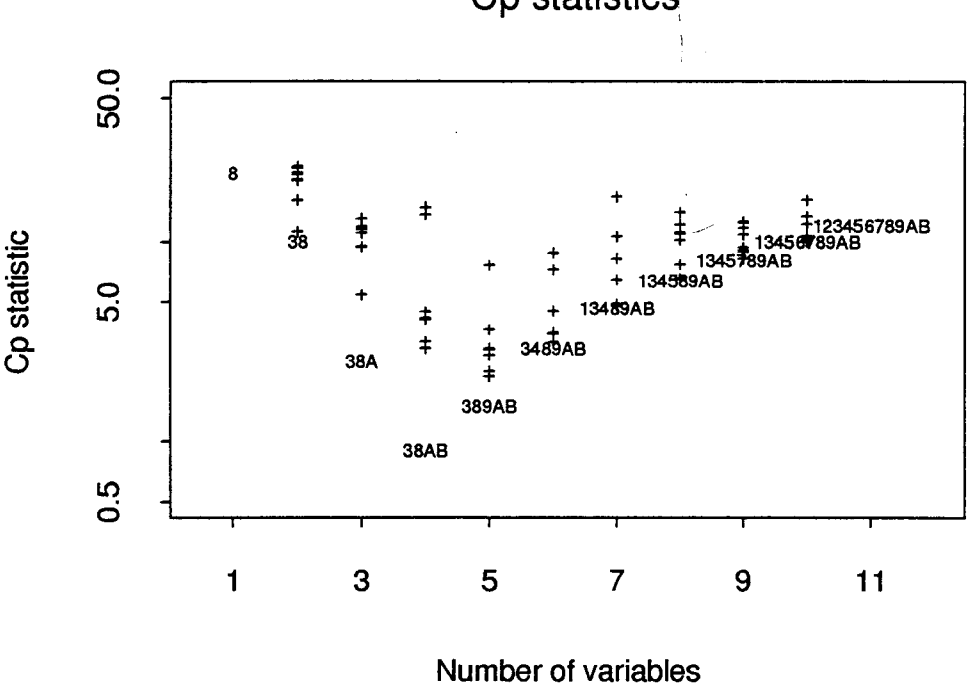
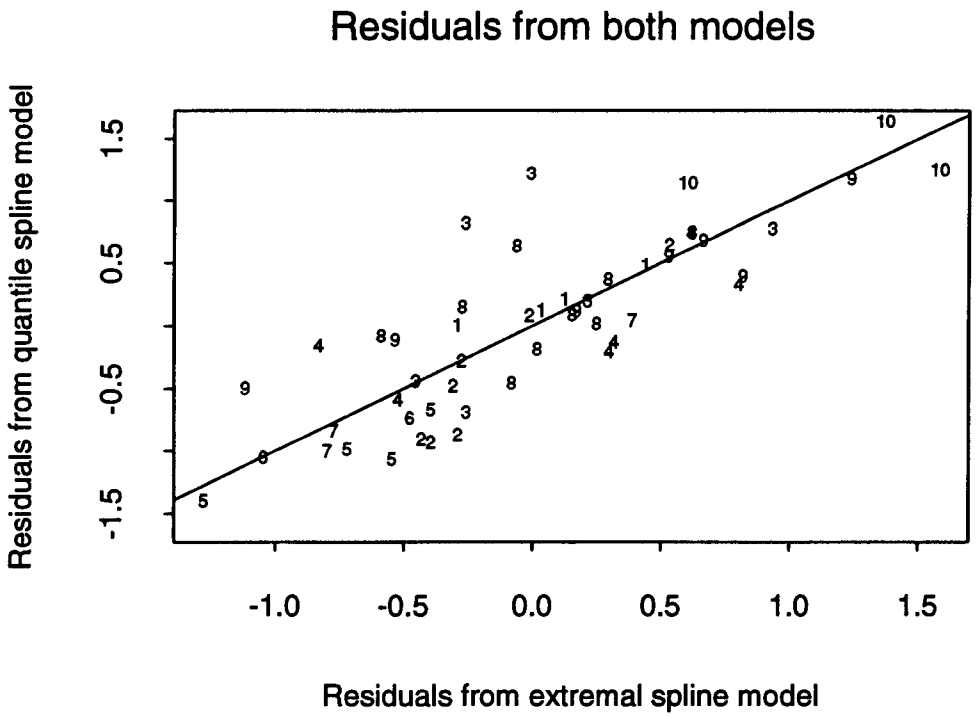


Figure 5. Residual plots for linear models



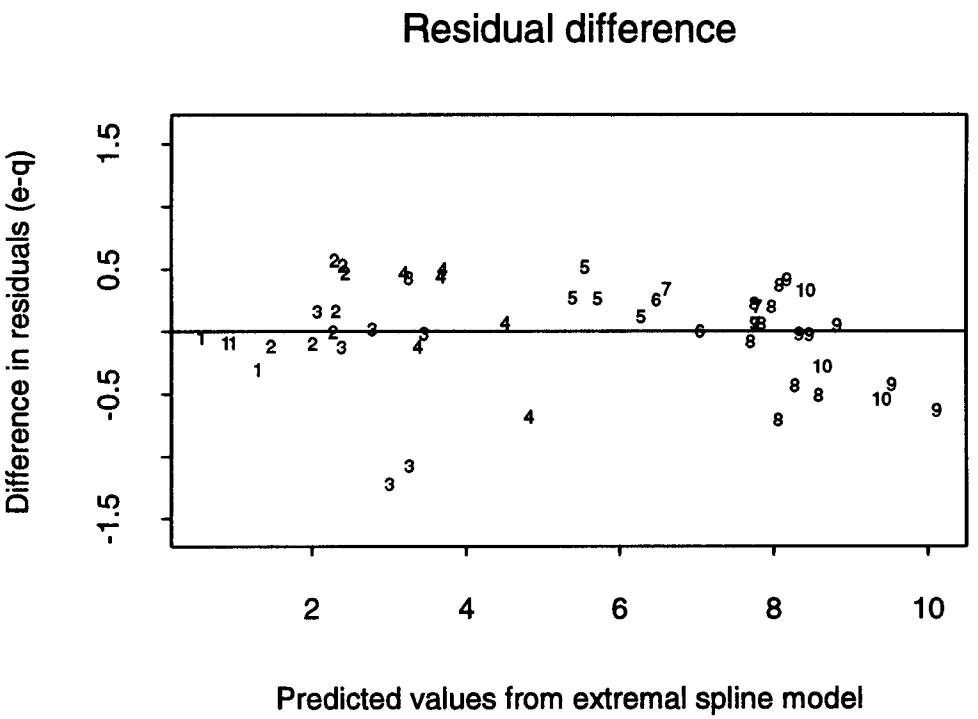


Figure 6. Comparison of extremal and quantile spline estimates of a grade 3 syringe

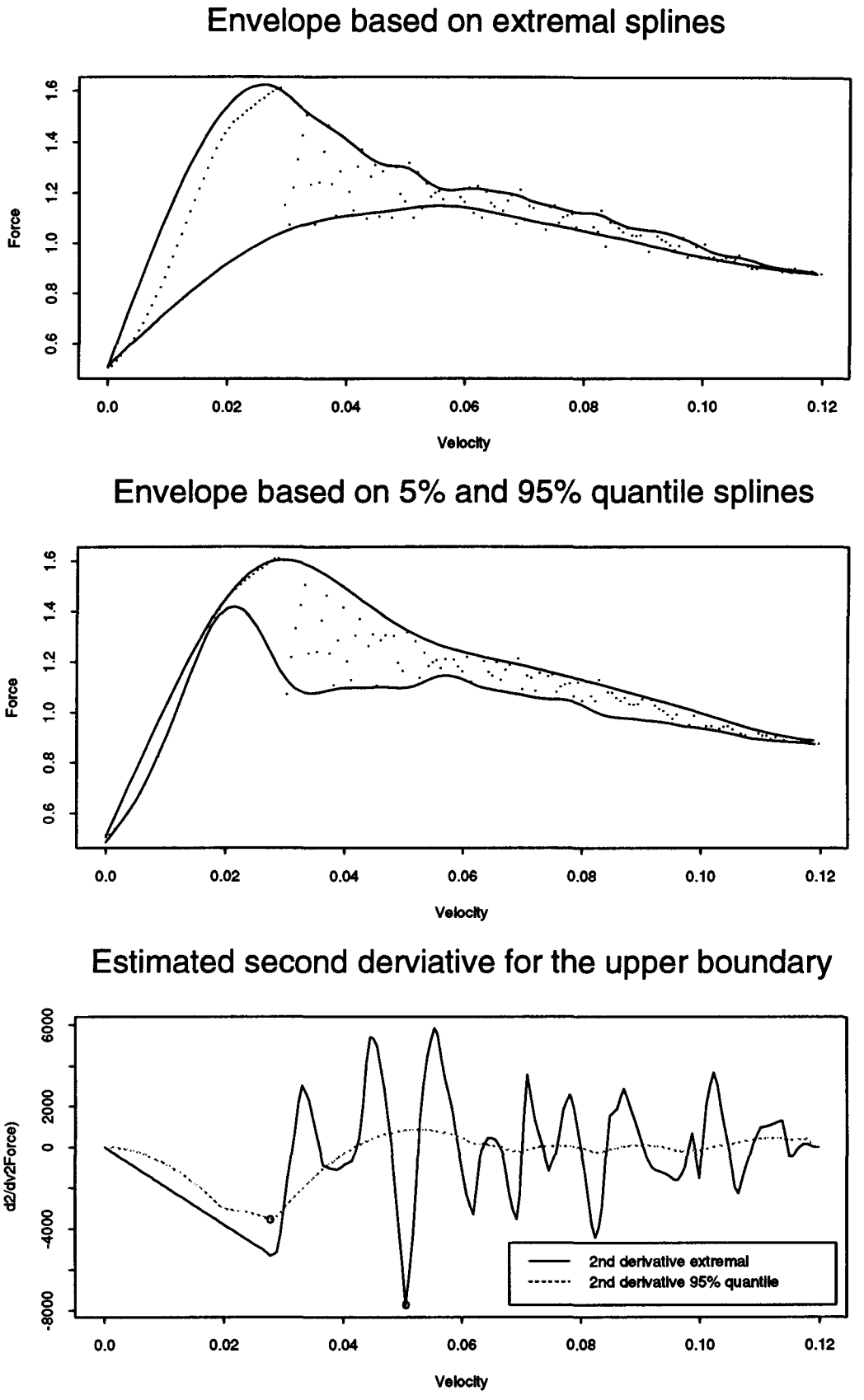
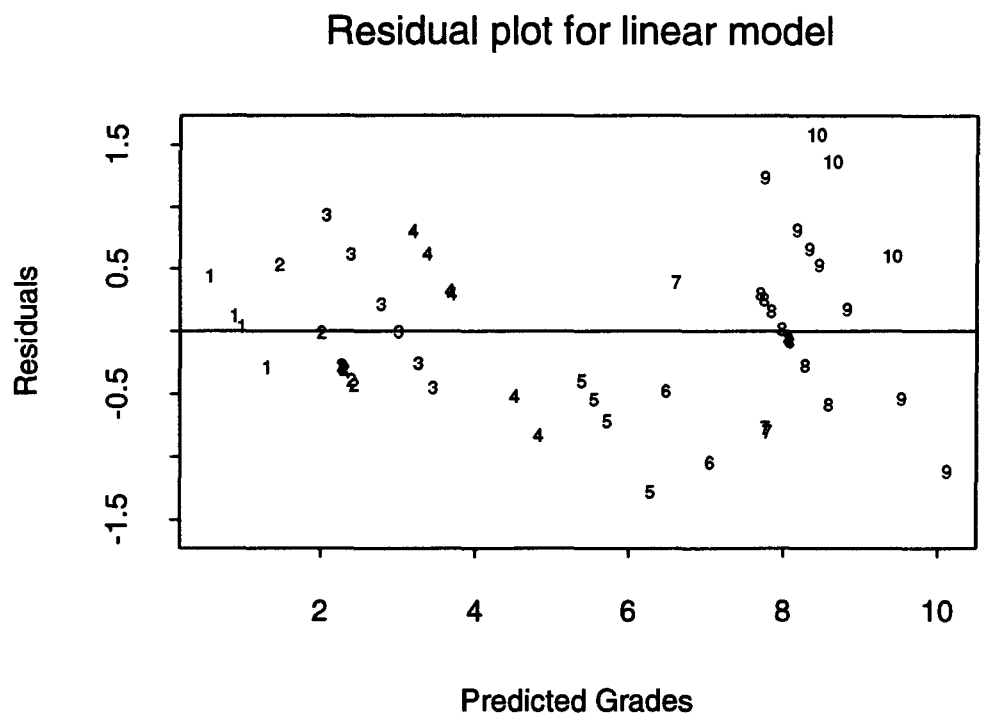


Figure 7. Non-linear classification model construction: features from extremal splines



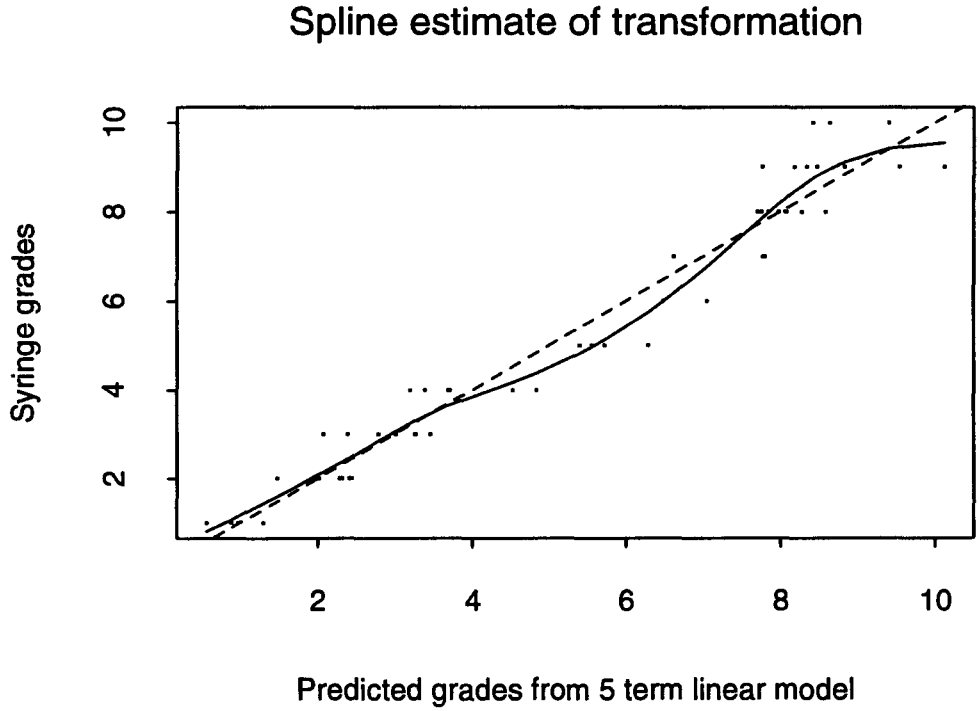
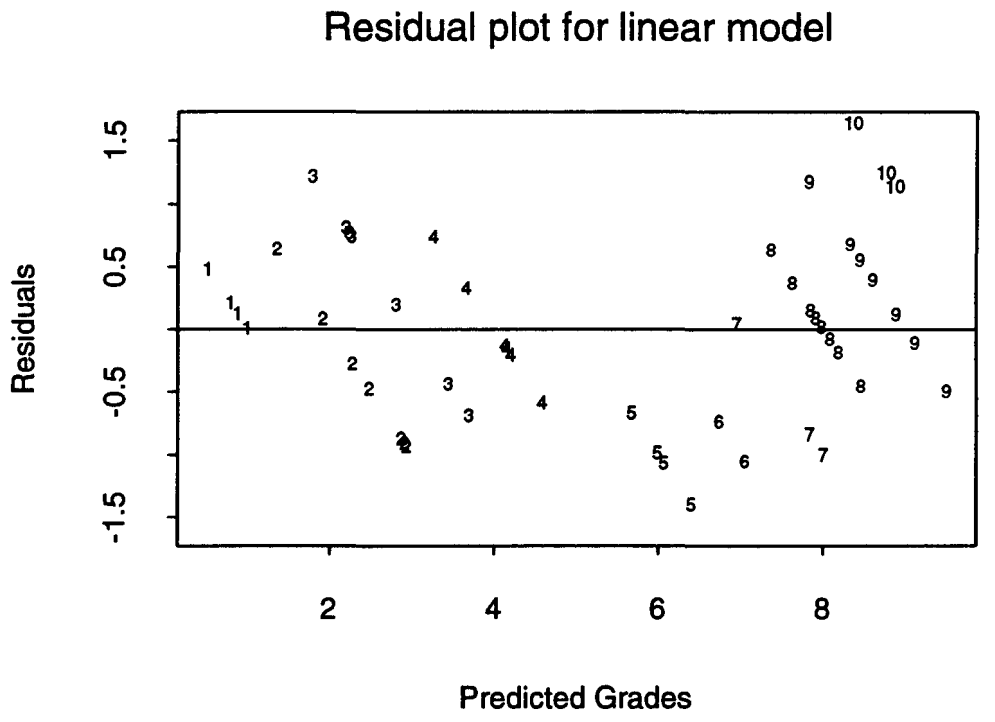


Figure 8. Non-linear classification model construction: features from quantile splines



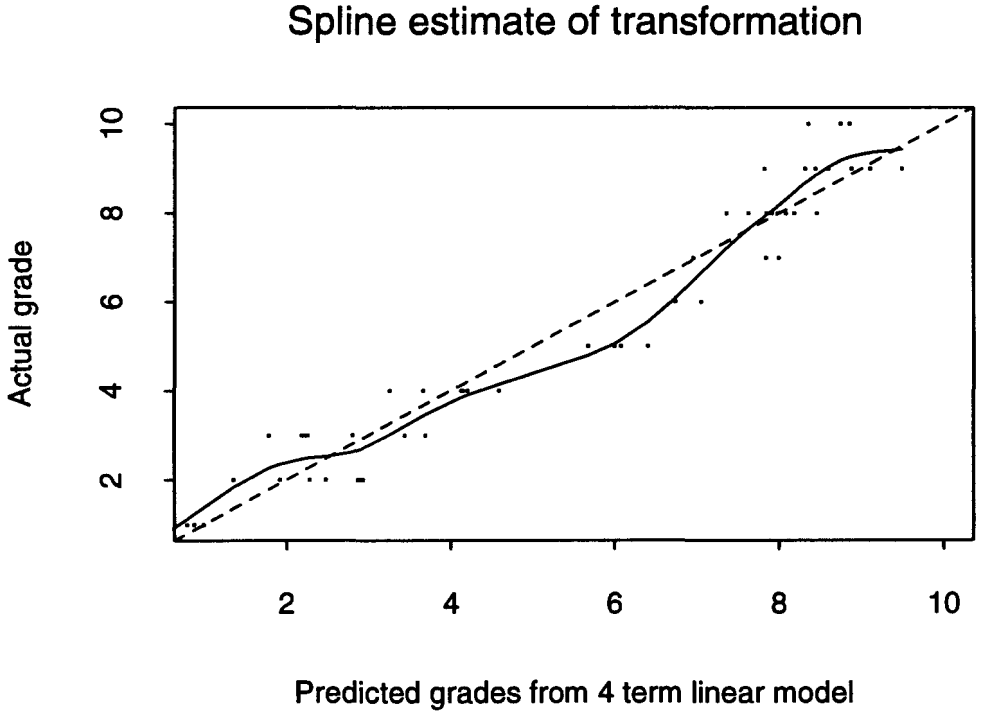
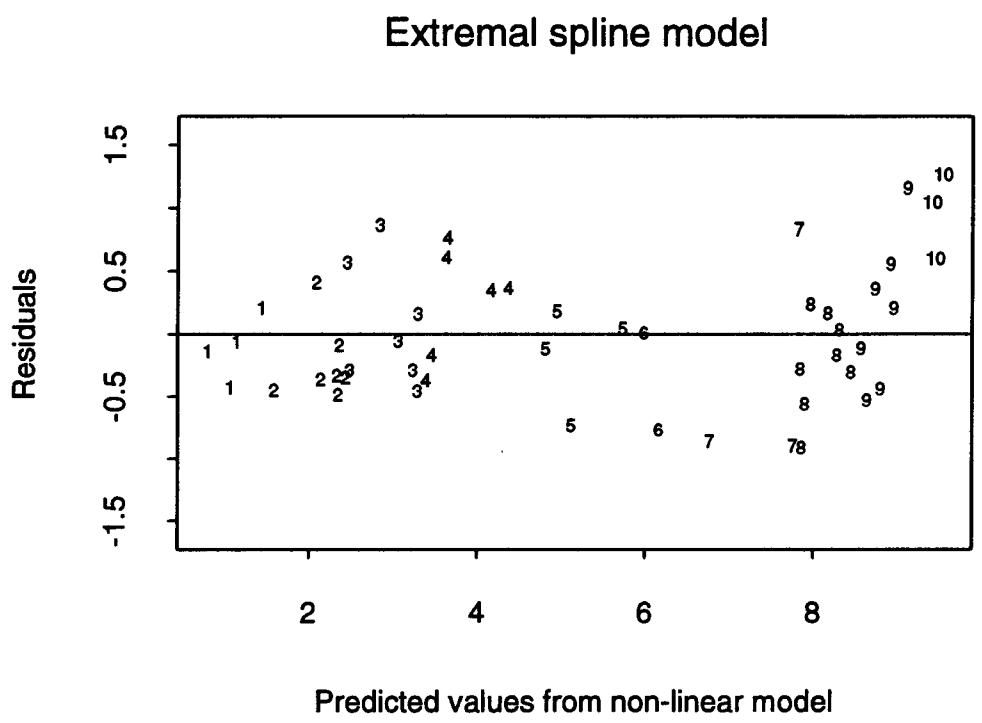


Figure 9. Residual plots for non-linear models



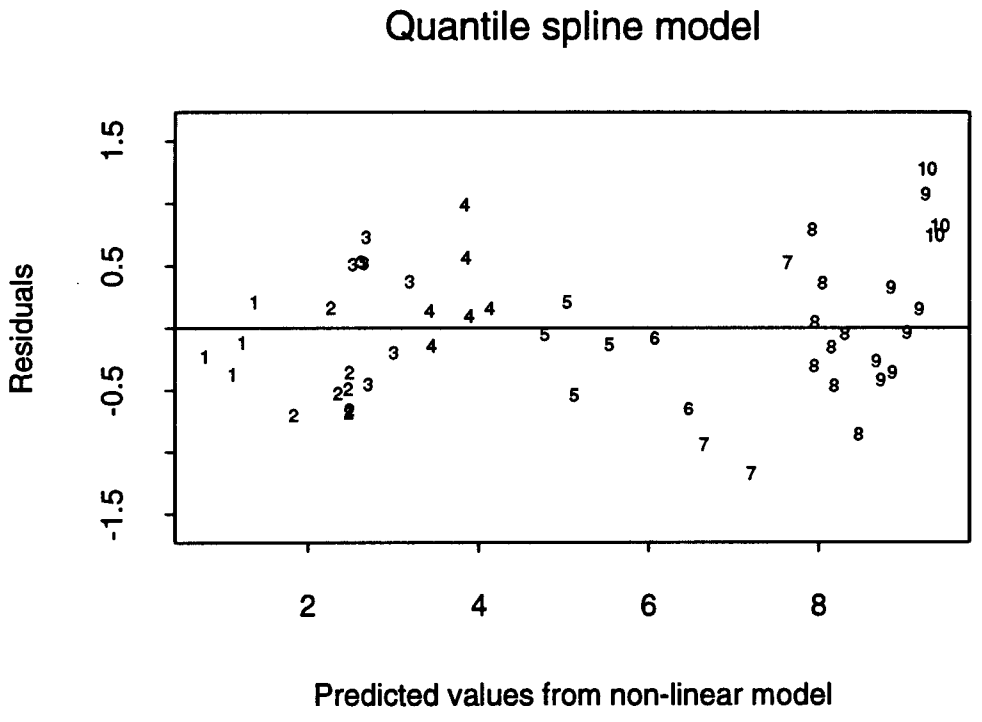


Figure 10: Cross-validated residual plots for non-linear models