A Nonparametric Regression Approach to Syringe Grading for Quality Improvement

Doug NYCHKA, Gerry GRAY, Perry HAALAND, David MARTIN, and Michael O'CONNELL

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 that incorporates the knowledge of the expert. The procedure is illustrated with the problem of 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 expert inspection. Using leave-one-out cross-validation, the prediction accuracy of the classification model is found to be about the same as the accuracy obtained from the expert.

KEY WORDS: Cross-validation; Model selection; Quantile splines.

1. INTRODUCTION

The classification of an individual production unit or clinical specimen based on a continuous 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 used in the classification. Such continuous recordings must be reduced to a useful summary for an appropriate diagnosis to be made. Often this is done subjectively by experts in the field of application. As an alternative, we suggest building a classification model that incorporates the experts knowledge. The key to this process is using nonparametric curve-fitting methods to extract useful features related to the expert's appraisal.

The approach is illustrated by the quality assessment of hypodermic diabetic syringes. These syringes are constructed with a rubber-stopped piston sliding within a plastic cylinder, aided by a lubricating fluid. Syringe quality is indicated by ease of use: A small force is required to actuate the syringe, and gradual pressure can be used to control the subsequent movement. 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 was achieved by modifying a standard testing instrument used in various engineering and scientific applications. The basic data from the instrument is a force/velocity curve, which we call a friction profile. The friction profile consists of several hundred measurements of force/velocity over less than a minute. Figure 1 gives some examples of friction profiles for three syringes of different qualities.

Although the testing procedure is not a realistic simulation of syringe use, an engineer familiar with the testing procedure and the hydrodynamic principles that control the frictional forces in this system can reliably grade the quality of a syringe. In the current application syringes are graded on a 10-point scale, with 1 the poorest syringe and 10 the best syringe. Based on careful examination of the force velocity curves, an expert can convert the friction profiles of syringes to grades with an accuracy of about ± 1–2 grades. The drawback of this type of procedure is that a specially trained individual is required, and even so there is still some subjectivity based on the visual assessment of friction profiles. Thus the test is not generally practical for quality improvement in a manufacturing setting and also may not 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. For each syringe, the friction profile is first summarized by smooth nonparametric curve estimates describing the mean and the upper and lower boundaries. From these, a smaller subset of features relating to the curves and their derivatives is extracted. A classification regression model is then built to predict the tribologists grade from the extracted features. Cross-validation is used to assess the predictive power of the model. For a diverse training sample of 51 syringes, the statistical classification approach is found to be accurate to within about one or two grades with 95% confidence.

The remainder of the article loosely follows the traditional scientific format of material, methods, results, and discussion. Section 2 describes the syringe data, Section 3 describes the feature extraction process based on nonparametric regression, and Section 4 describes the classification using regression models and the estimates of predictive power based on cross-validation. Finally, Section 5 provides some conclusions and discussion of alternative methods and problems.

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. Perry Haaland is Research Fellow, David Martin is Physicist, and Michael O'Connell is Scientist II, Statistics, Becton Dickinson Research Center, Research Triangle Park, NC 27709.

Journal of the American Statistical Association
December 1995, Vol. 90, No. 432, Applications & Case Studies

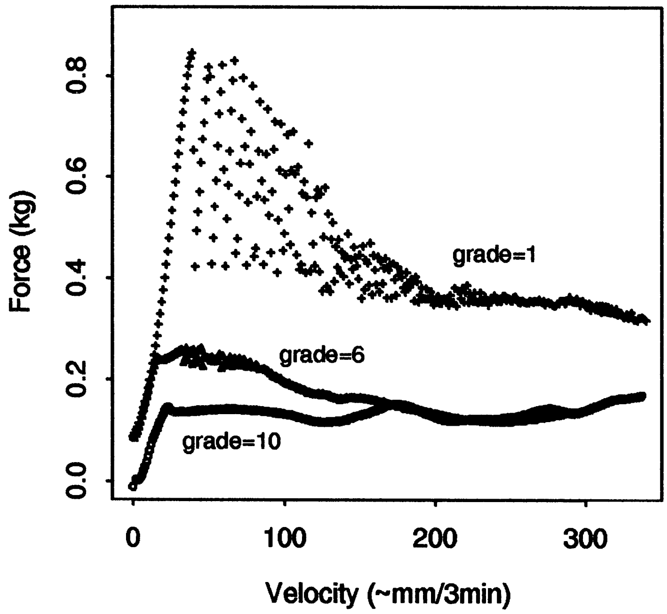

Figure 1. Examples of Friction Profiles From Testing Three Syringes. Friction profiles for grade 1 (+), grade 6 (▲), and grade 10 (○) syringes. Force is force required to maintain constant acceleration and is plotted against measured velocity.

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. 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. As the speed is increased, the system makes a transition to an indeterminant state between boundary and hydrodynamic lubrication. During this period, the syringe system oscillates between the high and low frictional forces associated with the boundary and hydrodynamic regimes, resulting in a phenomenon known as "stick-slip" or "chatter."

The testing 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.

Factors considered in the visual scoring procedure are listed in Table 1. Good syringes have friction profiles, $(Y = f(v))$, which are flat, smooth, and less than about 300 g. This is due to the absence of excessive boundary lubrication friction and stick-slip oscillations (ω = frequency of = 0 for all v). Poor syringes exhibit high forces in overcoming initial boundary friction changing to low force at high velocities, and frequently have stick-slip oscillations ($\omega \neq 0$). We refer to the initial force required to overcome boundary friction as the 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 more than 800 g. As the velocity is increased and the initial force 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 constant 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 a much lesser extent. The break-out force is much lower than that observed for profile 1, but is 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 those seen in profile 1. This profile is typical of some commercially available syringes.

Profile 10 (grade = 10) is an exceptional 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 to 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 the three 1 cc diabetic syringes given in Figure 1 were used to develop the feature extraction process. 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 in quality improvement applications. This range of syringe quality was achieved by varying the lubricant and treatment conditions in production of the individual syringes.

3. ESTIMATION OF DISTINCTIVE FEATURES

The most important step in the analysis is to reduce the friction profile data (approximately 300 points) to a lower-dimensional set of distinctive features useful for grading.

Table 1. Factors Influencing Visual Scoring

High grade	*Low grade*
$Y < 300$ g	$Y > 500$ g Y (low v) $\gg$ Y (high v)
$\omega = 0$, for all v	$\omega \neq 0$, some v $\omega \neq 0$, wide range of v

Y = frictional force
v = velocity of plunger
ω = frequency of oscillations

The goal in this part of the study is to reliably extract functionals of the friction profiles related to the qualitative features used to grade syringes. The functionals are derived from nonparametric curve estimates of the lower and upper boundaries of the friction profiles, f_L and f_U, that provide a smooth envelope for the measured force. The intent was for the upper boundary curve, f_U, to be used to capture 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. By experimenting with the friction profiles shown in Figure 1, we developed nonparametric procedures independently of the training sample of 51 syringes.

3.1 Spline Estimates for the Friction Profiles

The curve estimates considered in this work are all variations on a cubic smoothing spline where the smoothing is chosen adaptively from the data. It should be emphasized that these are being used solely for feature extraction and may not imply a valid statistical model. But for estimating the mean force, consider the nominal regression model,

$$Y_k = f(v_k) + e_k \qquad 1 \le k \le N,$$

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. 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 weight w_k may be specified as available, often as the reciprocal of some estimate of the variance of e_k.

The spline estimate at a particular point can be interpreted simply as a weighted local average of Y_k, $1 \le k \le N$, where the weights depend on the independent variable v_k, $1 \le k \le N$ and the smoothing parameter λ. In between observation points, the estimate has the form of a cubic polynomial, and thus it is simple to evaluate the estimate and its derivatives at arbitrary points.

The choice of the smoothing parameter, λ, is important. Because 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 w_k, $1 \le k \le N$, the spline estimate is a linear function of Y_k, $1 \le k \le N$. If $\hat{f}_\lambda$ denotes the spline estimate, then there is a "hat" matrix $\mathbf{A}(\lambda)$ such that $(\hat{f}_\lambda(v_1), \ldots, \hat{f}_\lambda(v_N))^T = \mathbf{A}(\lambda)Y$. With this notation, the generalized cross-validation (GCV) function is

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

where $\mathrm{RSS}(\lambda)$ is the residual sum of squares as a function of the smoothing parameter and $m(\lambda) = \mathrm{tr}\mathbf{A}(\lambda)$ 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, Green and Silverman 1994, and Hastie and Tibshirani 1990, chap. 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 call 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 minima. The second strategy, which estimates the 5% and 95% conditional quantiles of the friction profile data, is referred to as quantile splines.

To estimate an envelope that contains the limits of the force applied to the syringe, we take advantage of the physical properties of the data. 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. If it were 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, and one method for calculating such an envelope would be to find points that are local maxima and minima, then fit smooth curves through these points. A simple definition of a local maximum (minimum) is a point that is larger (smaller) than each of its two nearest neighbors. That is, based on the original velocity/force pairs, we form the subsets

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

and

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

Upper and lower envelope curves are then 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. 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.

Another approach is based on the fact that the instrument recordings are undersampled and thus do not provide exact information about the envelope of the friction profile. If one considers the traces in Figure 1 as regression data, then the range of the observations at different velocities can be recovered by estimating the conditional distribution. Specif-

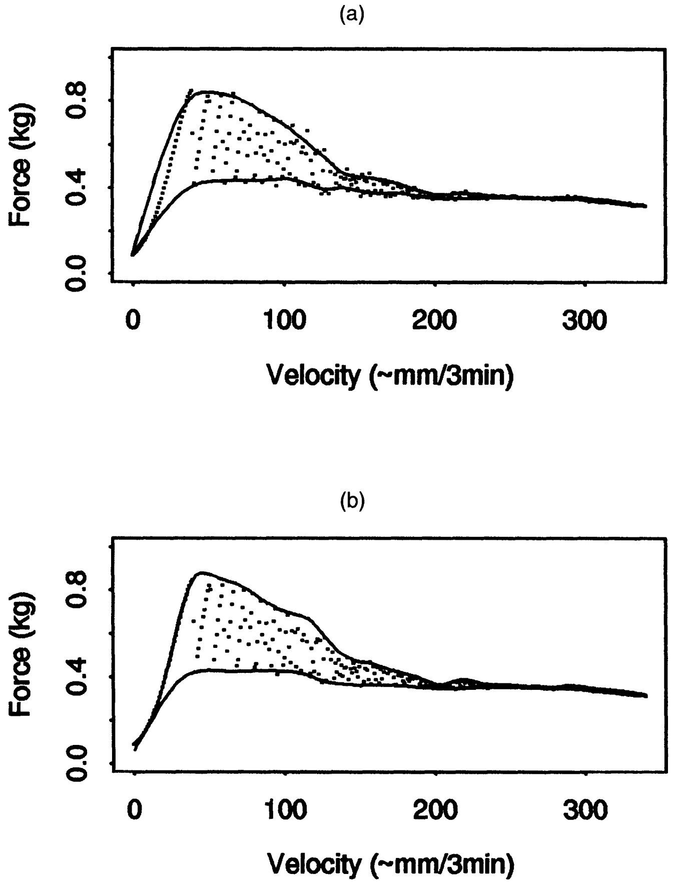

Figure 2. Extremal and Quantile Spline Fits to a Grade 1 Friction Profile, (a) Extremal Splines, (b) 5% and 95% Quantile Splines. Upper and lower envelopes are estimated using extremal and quantile splines.

ically, estimates of the 5% and 95% conditional quantiles might provide lower and upper bounds on the data at a particular velocity and considering these curves as a function of time yields estimates of the envelope. Based on the physical principles of the force measurements discussed earlier, a traditional regression model for these data is *not* correct, and thus one should be careful in interpreting the quantile estimates. But for extracting a smooth envelope from the friction profiles to provide features for classification, this is a reasonable strategy.

In a regression context, the conditional quantiles can be estimated using quantile splines (Bloomfield and Steiger 1983). These splines allow estimation of extreme quantiles of the data and thus capture some form of smooth boundary. Let

$$\begin{aligned}\rho_\alpha(u) &= (1-\alpha)|u| \quad \text{if } u \leq 0,\\ &= \alpha|u| \quad \text{if } u > 0.\end{aligned}$$

For any vector $\mathbf{Z}$, it well known that the αth 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. Because the least squares spline can be calculated rapidly, the computational burden for determining a quantile spline is not large. The appendix contains a more detailed description of this method. (The reader is referred to Gu 1992 and O'Sullivan et al. 1986 for more background on nonlinear spline smoothing methods.)

In view of the fact that this estimate is being used as a data-reduction mechanism and the regression model is not strictly valid, one needs to be careful about how the smoothing parameter is estimated. Usual data-based methods may not be appropriate for an undersampled oscillatory function. One conservative choice is to fix the roughness penalty at some particular value and then search on λ to find the quantile estimate with this roughness penalty. Another approach is to use cross-validation with a loss function based on the check function, ρ_α. Details of this method are outlined in the Appendix. Under this cross-validation scenario, visually accurate summaries of the envelope are obtained. Also, this method has the advantage of adapting to different amounts of smoothness for different syringe grades. One must keep in mind, however, that the basic model from which the quantile spline is derived is not strictly appropriate. So although the cross-validation approach works well for the data in hand, more closely spaced observations or less-rapid oscillations may lead to quantile estimates that interpolate the force curve rather than estimate the envelope.

Figure 2 also shows the 5% and 95% quantile spline estimates for the grade 1 friction profile data, where λ is found by the approximate cross-validation procedure mentioned earlier. Note that in contrast to the extremal 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, whereas the extremal splines are restricted to a subset (either local maxima or minima).

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.

3.3 Extracting Distinctive Features from the Friction Profiles

Based on the visual features used to grade a syringe, we considered the functionals listed in Table 2. The first functional is an overall measure of the complexity of the estimated upper boundary curve. The next two functionals are measures of the difference between the upper and lower curves and are intended to quantify the stick-slip effect. Two strategies are used to identify the break-out point: to find the point of maximum curvature in the upper envelope (4–6), and to identify the first local maximum in the profile (7 and 8). The last three measures attempt to summarize the smoothness of the profile after the break-out point 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 break-out point. Smoothness in this region may be characterized by the scatter in the first derivative of the upper curve, whereas a gradual trend in the profile is indicated by a positive median first derivative.

4. CLASSIFICATION OF SYRINGES FROM A TRAINING SAMPLE

For each of the 51 syringes tested, upper and lower boundary curves are estimated using extremal and quantile splines. Based on these curves, the 11 summary functionals are computed. The goal is then to build a model from these summary functionals that predicts the expert tribologist's syringe grade and hence enables the quality improvement engineer to classify syringes on a routine basis.

4.1 Subset Selection Among Features

We first consider models of the form

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

where Z_k is the grade of the kth syringe, $k = 1, 51$, as assigned by the tribologist; $X_{[k]}$ 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.

Because the main interest is in the predictive power of these summary measures, Mallows's Cp statistic is used to identify a subset of the functionals. Figure 3 presents a plot of the Cp statistics for the different subsets of features based on extremal splines. This model selection criterion indicates that a linear model based on eight features is the best choice for the extremal splines approach. 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 break-out (A).

Besides using Mallows's criterion, GCV is also considered. For linear regression, the GCV function for a subset of size p is $V(p) = (1/n)\mathrm{RSS}(p)/(1 - p/n)^2$. Under the assumption that p/n is small and $\mathrm{RSS}(p)/(n - p)$ is approximately equal to $\hat{\sigma}^2$ based on the full set of variables,

Table 2. Features of the Friction Profile Useful for Classification

1.	Effective number of parameters for $\hat{f}_U$, where the smoothing parameter is chosen by cross-validation as defined in Section 3.1
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 $\hat{f}_U$ (break-out point), V_{BP}
8.	Force at first maximum of $\hat{f}_U$, F_{BP} (break-out force)
9.	25th percentile of $\{\hat{f}_U'(v_k)\}$ for $v_k \geq V_{BP}$
A.	50th percentile of $\{\hat{f}_U'(v_k)\}$ for $v_k \geq V_{BP}$
B.	75th percentile of $\{\hat{f}_U'(v_k)\}$ for $v_k \geq V_{BP}$

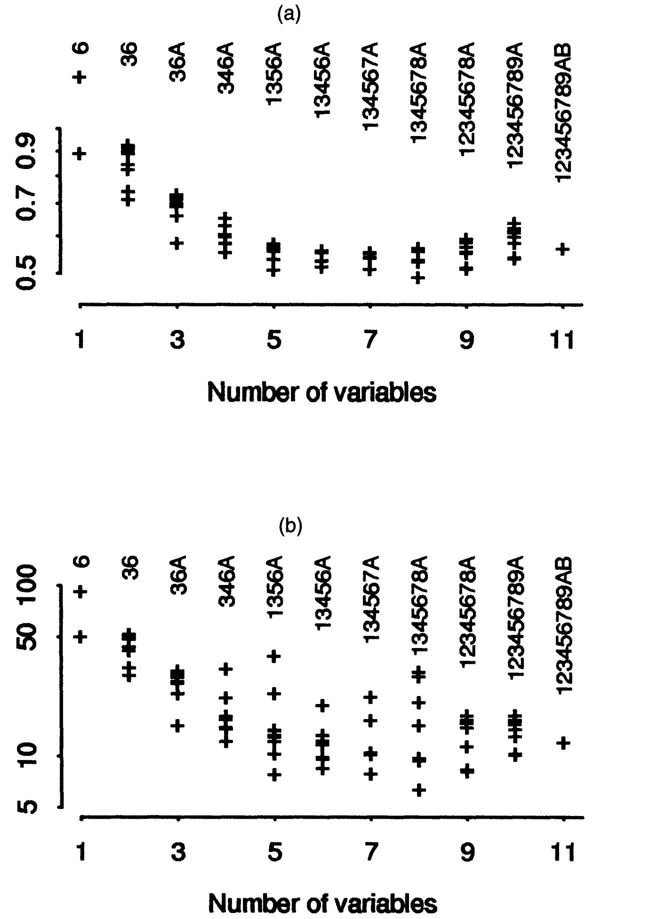

Figure 3. Regression Subset Selection Based on Features From Extremal Splines. The two selection statistics, (a) GCV and (b) Cp, are plotted against subset size. The text strings located at each subset size identify the subset with the smallest statistic. (See Section 3.3 for the feature codes.) Local minima of the model selection criteria occur for subsets of five and eight variables.

$V(p)$ is similar to $\hat{\sigma}^2$Cp. But when p/n is large, GCV is more conservative and penalizes larger models more heavily. For this reason, the GCV criterion is 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 (Fig. 3) indicate a local minimum (GCV $\doteq$ .5) at five features. The additional two features to the three described earlier 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 subset of five features is considered in subsequent analysis. The best subset of features based on quantile splines has four members—3, 8, A, and B—and has a slightly higher value for its GCV minimum.

4.2 Bias and Other Problems in Automatic Subset Selection

There are some concerns about subset selection without an independent data set for estimating the prediction error (Cook and Picard 1984). One important issue regarding using Cp and GCV for model selection is the tendency of

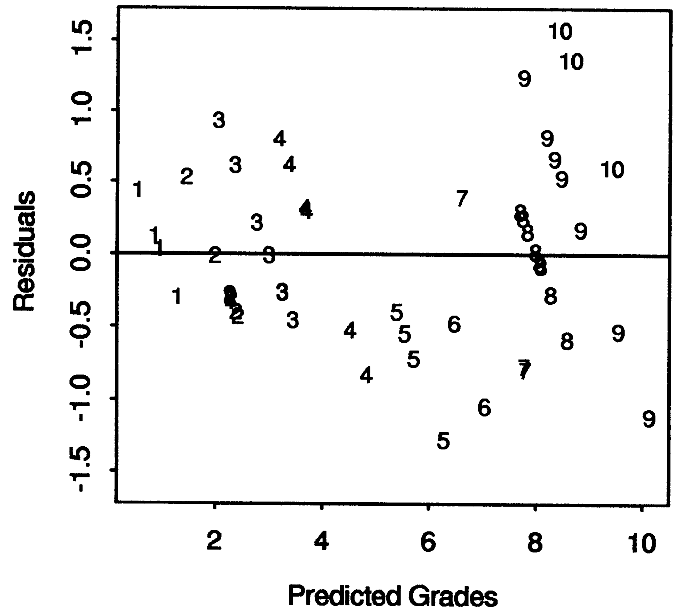

Figure 4. Residual Plot for a Linear Model With the Best Five Feature Subset as Obtained via Extremal Splines. Note some problems with the fit to intermediate grades.

these selection methods to include extraneous variables in the optimal subset. In fact, even in the limit as the sample size approaches infinity, there is a nonzero probability that additional variables are part of the estimated subset (Shao 1993). But in this application, it should be noted that the subsets used for our analysis are not particularly large, consisting of only 4 or 5 features, and are sensible, based on the expert's visual criteria for good syringes. The asymptotic results for GCV also indicate that one will not be led to a subset that is too small. In the limit the probability that all the significant variables are in the selected subset approaches 1. For this classification study, we believe that including all important features is more important than producing a model that has a minimal number of parameters. Thus for our goals, using the variable selection procedures is a reasonable strategy for identifying a useful subset.

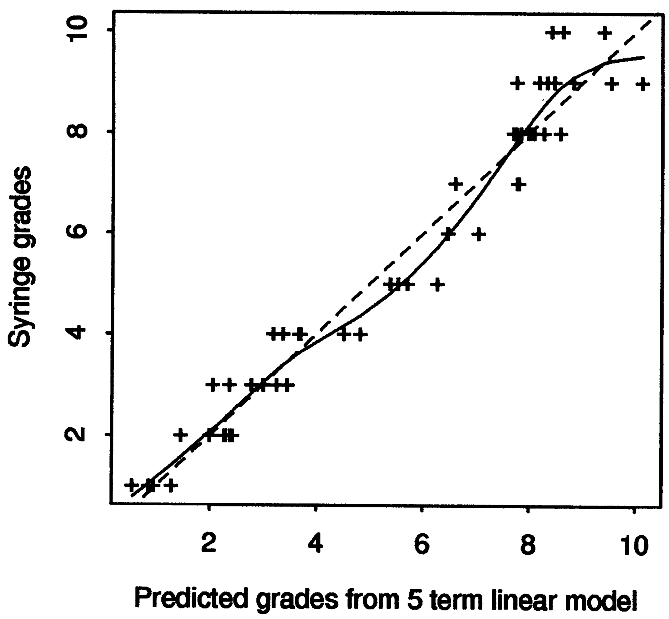

Figure 5. Estimated Spline Transformation of the Linear Predictions. The spline estimate of the transformation is obtained from a spline fit to the pairs ($\hat{\mathbf{Z}}_k$, $\mathbf{Z}_k$), where $\hat{\mathbf{Z}}_k$ are the predicted grades based on the five-term linear model of features obtained from extremal spline fits to the friction profiles.

Besides identifying the correct subset, another issue is whether the GCV criterion gives a reliable estimate of the prediction error. Such an estimate is consistent in the case of a fixed number of variables and an increasing sample size. But for finite sample sizes, the issue remains that the prediction error selected as a minimum over all subsets must necessarily underestimate the true prediction error. A conservative approach to this problem is to think of the subset selection as an integral part of estimating the model and to include it in a cross-validation procedure. The cross-validated residuals formed from such a procedure are useful for model checking, and the mean square of these residuals is a conservative estimate of the mean squared prediction error. Using this process, prediction error estimates of $(.81)^2$ for the extremal spline features and $(.86)^2$ for the quantile spline features are obtained. This represents an inflation of approximately 20% compared to the usual GCV estimate of the error.

4.3 A Nonlinear Classification Model

For all of the linear models based on subset selection, the residuals indicate some systematic departures from a linear function. The residual plots for the best five feature subset extremal spline model (Fig. 4) and the best quantile spline model both suggested that agreement with the actual grades can be improved if the estimated grades are transformed.

We thus consider the nonlinear model

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

where Z_k is the grade, $(X_{k1}, \ldots, X_{kp})$ are the functionals estimated from the kth friction profile, and H is a smooth transformation. A direct way to estimate 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 5 and 6 summarize the results of estimating H in this way using a smoothing spline based on the extremal features. A similar fit was achieved using the predicted values from the linear model based on the quantile spline features. 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 cross-validation residuals for the nonlinear (transformed) models (Fig. 6) appear to have less dependence on the predicted values than the untransformed linear model residuals (Fig. 4), but there is still some bias in estimating very high grades. After estimating the transformation, the parameter vector $\boldsymbol{\beta}$ is reestimated using nonlinear least squares, with the transformation considered fixed. From a practical viewpoint, this refinement of the parameter estimates give little improvement.

For this training sample of 51 syringes, the estimates of the prediction standard errors based on cross-validation residuals, constructed under the entire modeling procedure

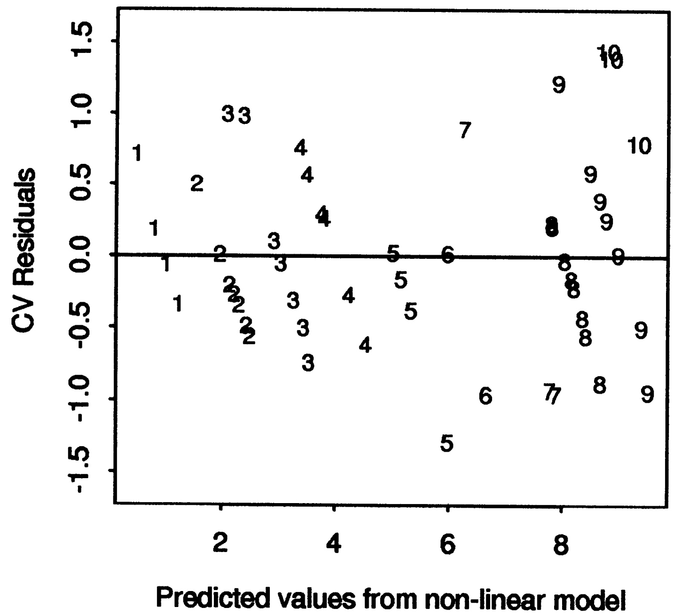

Figure 6. Cross-Validated Residuals From the Nonlinear Model. The residuals are obtained by successively omitting each data point and estimating the entire nonlinear classification model (including the model selection step in constructing the linear model). The residual is the difference between the omitted value and that predicted using the remaining data.

described earlier, are .76 for the extremal spline features model and .71 for the quantile spline features model on the transformed scale. Not surprisingly, these results are an improvement over assigning grades based on linear models. The transformed scale models thus can identify the syringe grade to within about ± 1.5 with about 95% confidence. One remaining systematic pattern in the residual plots is that the model does not do an adequate job in distinguishing between syringes of grades 9 and 10.

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

Other nonparametric regression models were also fit as potential classification models. These included a three-feature additive spline model identified by forward selection, an optimized multivariate adaptive regression spline (MARS) model, and a neutral net model identified by minimizing root mean squared error. These provided similar fits to the grading data, and all involved a subset of the features (3) mean difference between $\hat{f}_U$ and $\hat{f}_L$; (8) force at first maximum of $\hat{f}_U$; (A) 50th percentile of $\{\hat{f}'_U(v_k)\}$ for $v_k \geq V_{BP}$; and (B) 75th percentile of $\{\hat{f}'_U(v_k)\}$ for $v_k \geq V_{BP}$. Although these features are inherently interesting, they provided no further insight than the smooth transform nonlinear model and due to software considerations were not pursued for routine use in the quality improvement application.

5. DISCUSSION

The results from modeling the training sample of friction profiles indicate that syringes can be classified with an average accuracy of approximately ± 1.5 grades with 95% confidence. Because visual evaluation by an expert tribologist is only expected to be about this accurate, we believe that the approach makes a significant contribution to the automatic grading of syringes.

Our strategy for predicting syringe grades combines four basic ingredients: nonparametric 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 leave-one-out cross-validation of this model to estimate the average prediction error. Although some of the details are necessarily specific to the friction profile measurements, we believe that this work may serve as a blueprint for other studies involving prediction of semiqualitative properties from continuous recordings produced by an instrument. The approach also illustrates the potential for transferring knowledge from an expert to a routine user. 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.

APPENDIX: QUANTILE SPLINES

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

Background for the Computational Algorithm

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 (i.e., 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. Any value in between knots can be inferred from the piecewise cubic form. Thus an alternative version of the usual spline minimization problem is

$$\min_{\mathbf{h}\in\mathcal{R}^n} (1/N) \sum_{k=1}^{N} (Y_k - h_k)^2 w_k + \lambda \mathbf{h}^t \mathbf{R} \mathbf{h}, \tag{A.1}$$

where $h_k = h(t_k)$ and $\mathbf{R}$ is an $N \times N$ matrix that can be derived from the roughness penalty and the cubic spline basis.

The solution to (A.1) satisfies

$$2(Y_j - h_j)w_j + \lambda(\mathbf{R}\mathbf{h})_j = 0 \tag{A.2}$$

for $1 \leq j \leq N$.

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 (A.2). To carry this out, however, it is necessary to make a slight modification to ρ_α so that it is differentiable at zero. The idea is to round out the corner of ρ_α at zero by piecing in a quadratic function. Accordingly, let

$$\begin{aligned}\rho_{\alpha,\varepsilon}(u) &= \rho_\alpha \quad \text{for} \quad |u| > \varepsilon, \\ &= \alpha u^2/\varepsilon \quad \text{for} \quad 0 \le u \le \varepsilon, \\ &= (1-\alpha)u^2/\varepsilon \quad \text{for} \quad -\varepsilon \le u \le 0.\end{aligned}$$

In our calculations ε is chosen to be effectively zero relative to the magnitude of the data values.

If this approximation is applied to the minimization functional that follows, then one can characterize the solution in a similar manner to the weighted least squares spline in (A.1). Consider

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

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

$$\psi_{\alpha,\varepsilon}(Y_j - h_j) + \lambda(\mathbf{Rh})_j = 0 \qquad 1 \le j \le n, \qquad \text{(A.4)}$$

where $\psi_{\alpha,\varepsilon} = \rho'_{\alpha,\varepsilon}$.

Let $\mathbf{h}^0$ denote an initial starting value for the solution to (A.4). Rewrite the system as

$$2(Y_k - h_k)\left[\frac{\psi_{\alpha,\varepsilon}(Y_k - h_k)}{2(Y_k - h_k)}\right] + \lambda(\mathbf{Rh})_k = 0, \qquad \text{(A.5)}$$

then substitute the initial estimate for h_k in the bracketed term. With this substitution, one obtains a linear system of equations in $\mathbf{h}$. Identifying w_k with $[\psi_{\alpha,\varepsilon}(Y_k - h_k^0)/2(Y_k - h_k^0)]$, this approximate system is solved using the standard algorithm for weighted least squares smoothing splines. In general, if $\mathbf{h}^J$ is the vector at the Jth iteration, then $\mathbf{h}^{J+1}$ is obtained by solving the linear system where the weights are based on $\mathbf{h}^J$. Note by construction of the weighting function, at convergence $\mathbf{h}^\infty$ will solve (A.4).

In our computations the quantile splines are estimated for a grid of smoothing parameters and the estimates are computed in descending order with respect to λ. An ordinary least squares spline is 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. Because 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.

Approximate Cross-Validation

Let $\hat{f}^{[k]}_{\lambda,\alpha}$ denote the quantile spline having omitted the kth data point, (v_k, Y_k). Now suppose that $\hat{f}^{[k]}_{\lambda,\alpha}$ has been computed for each k. One may then consider the quantile cross-validation (QCV) criterion,

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

as a measure of the fit of the quantile spline to the data. But because the quantile spline is a nonlinear function of Y, QCV is computationally intensive to calculate, and we consider an approximation.

If $\mathbf{A}_w(\lambda)$ is the associated "hat" matrix for the quantile estimate at convergence, then one may consider the approximate leave-one-out estimate,

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

This form has an attractive interpretation in terms of the computational algorithm. Using the quantile spline for the full data as a 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 seems adequate to approximate the leave-one-out estimate for the purpose of constructing the comparative statistic

$$(1/N)\sum_{k=1}^{N}\rho_\alpha\left(\frac{Y_k - \hat{f}_\lambda(v_k)}{1 - \mathbf{A}_{w,kk}(\lambda)}\right).$$

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

[Received May 1994. Revised May 1995.]

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.

Breiman, 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.

Craven, P., and Wahba, G. (1979), "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerical Mathematics,* 31, 377–403.

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

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines" (with discussion), *The Annals of Statistics,* 19, 1–141.

Gu, C. (1992), "Cross-Validating Non-Gaussian Data," *Journal of Computational and Graphical Statistics,* 1, 169–179.

Green, P., and Silverman, B. (1994), *Nonparametric Regression and Generalized Linear Models,* New York: Chapman and Hall.

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

O'Sullivan, F., Yandell, B. S., and Raynor, W. J., Jr. (1986), "Automatic Smoothing of Regression Functions in Generalized Linear Models," *Journal of the American Statistical Association,* 81, 96–103.

Picard, R. R., and Cook, R. D. (1984), "Cross-Validation of Regression Models," *Journal of the American Statistical Association,* 79, 575–583.

Reinsch, C. (1967), "Smoothing by Spline Functions," *Numerical Mathematics,* 10, 177–183.

Shao, J. (1993), "Linear Model Selection by Cross-Validation," *Journal of the American Statistical Association,* 88, 486–494.