

Nonparametric Regression Analysis of Data from the Ames Mutagenicity Assay

by John B. Cologne¹ and Norman E. Breslow²

The Ames assay has received widespread attention from statisticians because of its popularity and importance to risk assessment. However, investigators have yet to routinely apply modern regression methods that have been available for more than a decade. We study yet another approach, the application of nonparametric regression techniques, not as the ultimate solution but rather as a framework within which to address some of the shortcomings of other methods. But nonparametric regression is itself prone to difficulties when applied to Ames assay data, as we show through the use of two examples and some simulation studies. We argue that there remains a great need for further development of statistical methods suitable to the Ames assay. It is hoped that such work can be stimulated and guided by greater collaboration between statisticians and laboratory investigators.

Introduction

The Ames (Salmonella/histidine reversion) mutagenicity assay (1,2) may be the most popular short-term procedure for the evaluation of chemical mutagenic and carcinogenic potency. Because of its central role in risk assessment, the assay has received the attention of numerous statisticians who have sought to develop and apply appropriate methods of analysis with the goal of providing valid tests of mutagenicity and estimates of potency. Breslow and Kaldor (3) reviewed the wide range of statistical methods that have been proposed for use with the Ames assay and other mutagenicity assays. In particular, modern methods involving regression models and incorporating assumptions concerning biological mechanisms and distributional properties of the response data have become available and should be preferred over the use of simpler, more classical methods.

In light of these advances, it might be expected that investigators would routinely apply modern methods of analysis. Yet, an informal scan of recent volumes of one major journal revealed that these methods are almost never used by investigators. Most inferences regarding mutagenicity seem to be based on classical methods of analysis (e.g., simple regression or ANOVA, multiple *t*-tests, etc.) or even nonstochastic methods such as the two-fold rule (4). The problem of bridging the gap between statisticians and toxicologists therefore remains to be overcome. This contrasts sharply with the fields of epidemiologic research and

clinical trials, where modern regression methods are much in vogue.

Given this state of affairs, is it reasonable to introduce new methods and to further develop existing ones? Should not efforts be directed toward facilitating better application of existing methods? We would argue that both problems require further attention, namely: *a*) Statisticians have not yet achieved an ultimate solution to the problem of how best to analyze Ames assay data; and *b*) investigators should make concerted efforts to understand and apply, with the aid of statistical collaborators, modern methods of analysis. In this way, investigators can help statisticians to evaluate the utility of various methods by providing input as to their performance on a large amount of real data. Such feedback should enable statisticians to develop better methods.

We should interject at this point that we strongly advocate collaboration between statisticians and laboratory investigators. We do not deem simplicity or ease of understanding on the part of nonstatisticians to be the only criterion for recommending methods of analyzing laboratory data. Proposing overly simple and statistically inefficient methods to make them palatable to statistically untrained researchers can lead to misleading results and, perhaps, distrust of statisticians. We hope that this conference, by bringing together statisticians and toxicologists, will allow toxicologists to acquire better understanding of statistical methods and help build the kind of collaboration that will further the interests of both fields of endeavor.

With these introductory ideas in mind, we turn now to focus specifically on the first issue, the need for further development of methods of analysis. We will not address the issue of facilitating better application. We propose a new direction for methodological development in the area of the Ames assay — nonparametric regression — and present a specific example based on the smoothing spline. This paper is not meant to be the final word on statistical methods for the analysis of Ames data, nor do we wish to imply that the present method is the only

¹Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732 Japan.

²Department of Biostatistics, SC-32, School of Public Health and Community Medicine, University of Washington, Seattle, WA 98195.

Address reprint requests to J. B. Cologne, Department of Statistics, Radiation Effects Research Foundation, 5-2 Hijiyama Park, Minami-ku, Hiroshima 732 Japan.

This paper was presented at the International Biostatistics Conference on the Study of Toxicology that was held May 13-25, 1991, in Tokyo, Japan.

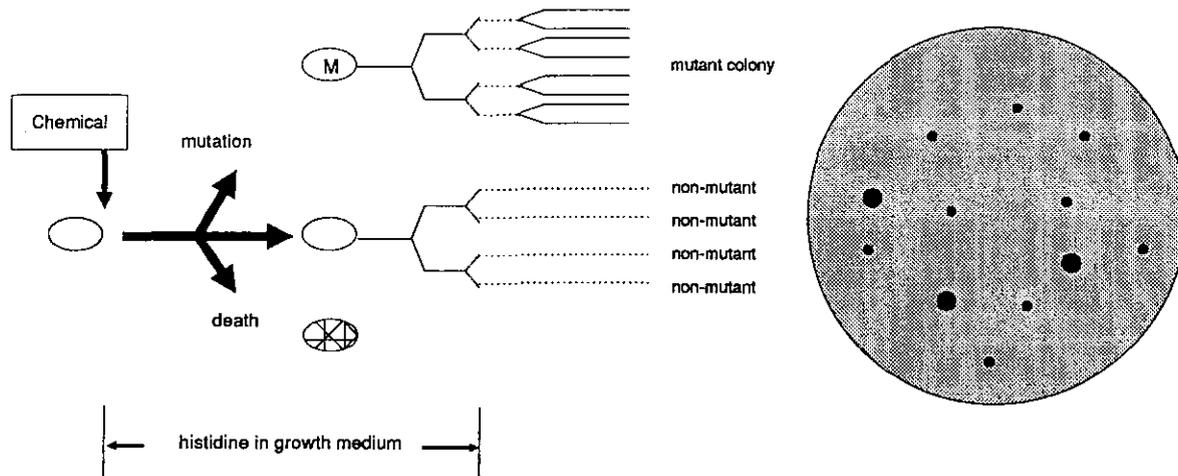


FIGURE 1. Illustration of the Ames assay. Mutant bacteria (*S. typhimurium*; ovals) are applied to plates containing growth medium (including histidine) and the test chemical. The chemical may cause mutation (M) or cell death, or it may have no effect. After histidine is depleted, only histidine-independent reverse mutants can continue to divide, producing colonies that are visible as spots on the plate (right). The microorganisms that survive but do not reverse-mutate form a discoloration of the plate called the "background lawn." With severe toxicity, most microorganisms are killed, so that the background lawn is absent; such plates are typically excluded from analysis.

possible application of nonparametric regression methods. In fact, it will be seen that this new method may at present have only limited usefulness. Rather, we use our approach as a framework within which to evaluate state-of-the-art methods and characterize the need for further research. In the end, we hope it will be obvious that there remains a great deal of work to be done to overcome problems that exist in currently available methods. This work could be in the form of better alternative methods, further development of existing methods, or comparative studies to provide guidance as to when and how existing methods should be applied. It is expected that all of these can and should be performed in close collaboration with laboratory investigators.

Overview of the Ames Assay

Figure 1 summarizes the Ames assay. Mutant strains of the bacterium *Salmonella typhimurium*, which are dependent on histidine for growth, are plated on petri dishes that contain the test chemical and a limited amount of histidine in the growth medium. The end point is reverse mutation to histidine independence, which is detected by the existence of surviving colonies of bacteria after histidine has been depleted. The outcome of interest for statistical analysis is the frequency of such colonies. Several doses of test chemical are used, including a zero-dose control. Each dose is typically tested in duplicate or triplicate. Interesting statistical features of data arising from the Ames assay include overdispersion (relative to Poisson sampling) and competing underlying dose responses (mutagenicity and toxicity).

It is often desirable to condense the results of an assay to a single number, the "mutagenic potency," to rank and compare different chemicals. The most common potency measure used in the Ames assay is the initial slope of the mutagenicity dose response. The two major goals of analysis are to test for mutagenicity and to estimate mutagenic potency (given a positive test). Often, the mutagenicity test is based on the potency estimator, but there are methods that test mutagenicity independently from

potency estimation [e.g., Simpson and Margolin (5)]. Throughout this paper, we use the initial slope both to estimate potency and to test for mutagenicity. We leave it to others to discuss potency measures and their relative advantages (6,7).

Two examples from the literature demonstrate the wide variety of data encountered in the Ames assay. The first (Fig. 2a)

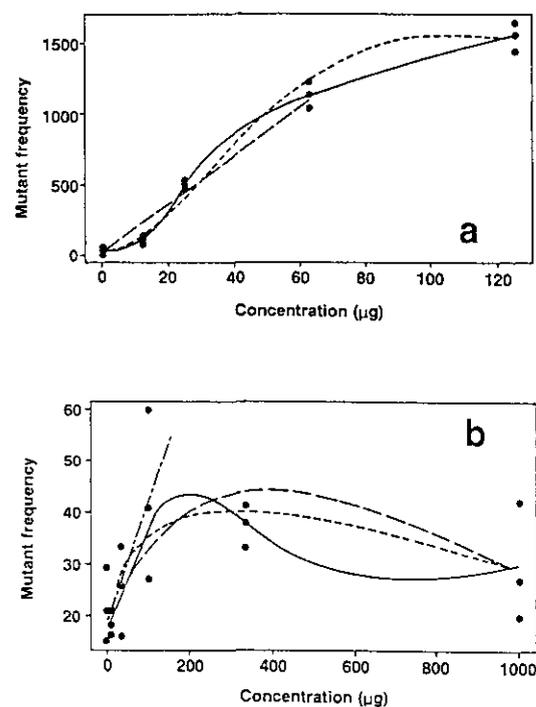


FIGURE 2. Two examples of Ames assay results from the literature. (a) Coal tar base extract data (8). Fitted curves represent point rejection (—), nonlinear regression using Equation 1 (---), and the smoothing spline (-.-). (b) Quinoline data (9). Fitted curves represent point rejection (---), nonlinear regression using Equation 1 without the quadratic term (—), quasi-likelihood with a loglinear model (-.-), and the smoothing spline (-.-.-).

concerns mutagenicity of base extract from crude tar effluent following coal gassification (8). In this example, the presence of a mutagenic effect seems clear, and we can ascertain the nature of the dose response from a plot of the data. Note, however, the possibility of little or no mutagenic effect at low doses because of the quadratic shape of the low-dose response. Fitted lines are the estimated dose responses from nonlinear regression, point rejection, and nonparametric regression (see subsequent sections).

The second example (Fig. 2b) is from a study of the mutagenicity of the chemical quinoline (9). In this example it is not easy to ascertain the dose response because of the variability in the data, although both mutagenicity and a downturn due to toxicity at high doses are evident. Fitted curves are from the point rejection, nonlinear regression, quasi-likelihood, and nonparametric regression methods (see subsequent sections).

We will keep mathematics to a minimum, focusing rather on concepts. Nevertheless, some basic notation is required. Let Y_{ij} be the observed frequency at dose i ($i = 1, \dots, K$), replicate j ($j = 1, \dots, n_i$) and let the actual value of the i th dose be x_i ($x_i = 0$). Let the mean and variance of Y_{ij} be μ_i and σ_i^2 , respectively.

The total sample size is $n = \sum_{i=1}^K n_i$. Finally, call the mutagenicity and survival functions $M(x)$ and $S(x)$, respectively, and let the composite dose response function be $\mu(x)$, the expected mutant frequency given dose x .

Regression Analysis of Ames Assay Data

Breslow and Kaldor (3) have reviewed statistical methods for analyzing Ames assay data. We mention here only modern methods that specifically accommodate regression analysis of Poisson data. All such methods are based in some way on the model of bacterial mutagenesis of Haynes and Eckardt (10),

$$\mu(x) = N_0 M(x) S(x)$$

where $M(x)$ and $S(x)$ are exponentiated polynomials and N_0 is the (unknown) number of plated organisms. A typical example is

$$\mu(x) = (m_0 + m_1 x + m_2 x^2) e^{-sx} \quad (1)$$

where $M(x)$ is approximated by its exponent, which is presumed to be small. [Note that N_0 has been absorbed into the parameters of $M(x)$; we subsequently write $M_M(x)$ in place of $N_0 M(x)$.] Such models have been fit by nonlinear regression (8), Poisson maximum likelihood (11), quasi-likelihood (12), and adaptive procedures [e.g., the point rejection method (13) based on the negative binomial likelihood]. The parameters m_0 (spontaneous mutant frequency) and m_1 (mutagenic potency) in Equation 1 may be constrained to be nonnegative if desired, but this is not necessary in practice. Background mutation frequency is typically large enough that m_0 is not estimated close to 0, and negative estimates of m_1 can be taken as no evidence of low-dose mutagenicity.

Margolin and others (9,14) have generalized these models to incorporate multiple generation effects in the assay, fitting their models with maximum likelihood using the negative binomial

distribution. In this paper we use their multigeneration model IV (9), with three generations for mutagenesis and histidine depletion and unlimited toxicity, to generate random data for evaluating the performance of various methods. The form of this model is

$$\mu_{3,\infty}(x) = N_0 M(x) \frac{[2S(x)-1]_+}{S(x)} \sum_{i=0}^2 (2[1-M(x)]S(x))^i \quad (2)$$

where $[a]_+ = \max(0, a)$.

Several restrictive assumptions accompany the use of Haynes-Eckardt models. One is that mutagenicity and toxicity are stochastically independent. If such independence does not hold (e.g., mutants have different survival rates), then a more general model may be needed. Another is the lack of metabolic and mutagenic mechanisms in the model. Indeed, one major criticism of the Ames assay as used for human cancer risk assessment is that it lacks the complex chemical and cellular mechanisms involved in human carcinogenesis [briefly reviewed by Weinstein (15)].

Apart from the issue of validity of the basic model itself, other difficulties arise in the case of specific methods. For example, the point rejection method circumvents toxic effects by rejecting high-dose points when there is a significant downward departure from linearity. Two potential sources of bias are: *a*) model misspecification when the dose response is nonlinear and *b*) one-sided (downward only) point rejection, which can occur by chance even in the absence of toxicity. The multiple generation model is theoretically quite attractive, but the choice of generation times can have a substantial effect on the magnitude of the estimated potency (9).

Although all methods have potential deficiencies in the response model, all the methods mentioned above accommodate nonconstant Poisson variance and overdispersion (with the exception of maximum Poisson likelihood). In practice, they seem to perform well in many examples, but more widespread use of these methods would facilitate a much-needed critical evaluation of their general applicability. In a later section, we use Monte Carlo studies to characterize, for some of these methods, circumstances in which they are most appropriate.

Nonparametric Regression

In this section we only briefly describe one approach to analyzing Ames data by nonparametric regression. Technical details will be made available elsewhere (Cologne, in preparation); a brief outline is presented in the Appendix.

Motivation

Before introducing our approach mathematically, it may be useful to motivate it from a nontechnical perspective. The hypothetical data shown in Figure 3 provide an exaggerated illustration of the concept of fitting nonparametric functions to data. Most investigators will be familiar with the use of the sample means plus or minus two standard errors as depicted in Figure 3a. A straight line connecting the means gives some indication of the dose response, but it is a crude and inefficient estimate. Furthermore, the initial slope estimate is based only on the first two dose levels, ignoring information about the shape of the dose

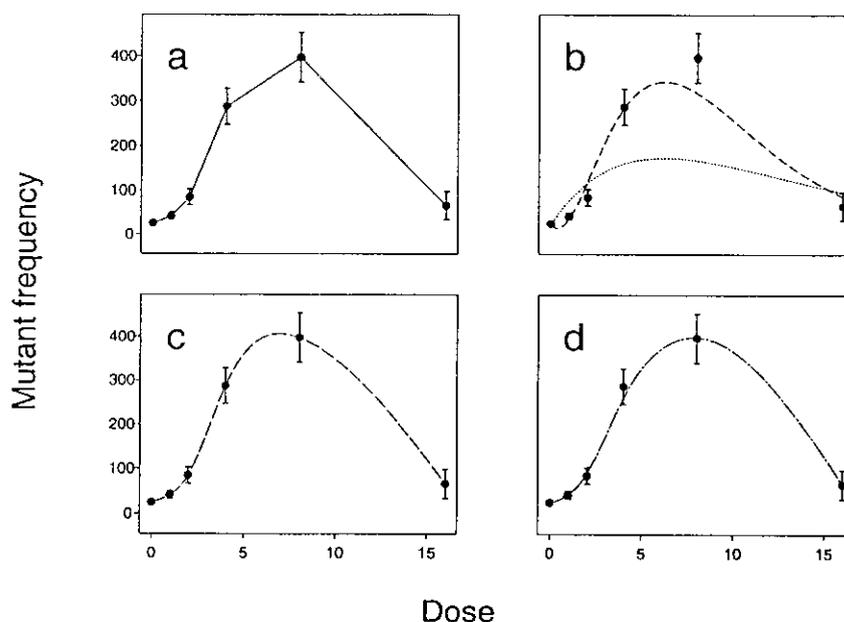


FIGURE 3. Hypothetical Ames assay data used to illustrate the concept of nonparametric regression. (a) Simple straight lines connecting sample means. (b) Haynes-Eckardt dose-response models with linear (\cdots) and linear-quadratic ($---$) mutagenicity (Equation 1). (c) A cubic interpolating spline. (d) A cubic smoothing spline. Panels a and c illustrate interpolating curves, which frequently overfit the data, producing highly variable potency estimates. Panels b and d represent various degrees of smoothing by a mathematical model; the cubic smoothing spline model is less restrictive than the Haynes-Eckardt dose-response model.

response. Finally, the model interpolates the sample means, thereby ignoring their variability.

Parametric models attempt to account for variability in the sample means by smoothing them according to some mathematical model (such as Equation 1 of the previous section). Figure 3b shows linear and linear-quadratic Haynes-Eckardt models fit to these data. Initial slope estimates based on these models incorporate information from the entire experimental range of doses, not just the first two doses. However, neither model fits the data well; presumably they do not reflect the true mechanistic model. In addition, different models with seemingly equivalent fits to the data can produce disparate potency estimates (as is well known to occur in the case of low-dose extrapolation in the animal bioassay).

As an alternative, we consider using a general class of mathematical functions, rather than the specific functions implied by use of the Haynes-Eckardt model. Figure 3c shows the fit of a "cubic interpolating spline," a piecewise polynomial function that connects points in a smooth fashion. (In fact, the line in Figure 3a is a linear interpolating spline.) The amount of smoothness at the dose values depends on the degrees of the polynomials (compare Figure 3c to Figure 3a). We can add higher-order polynomial terms to make the function more smooth at the dose values. Generally, cubic splines are sufficiently smooth so that higher-order terms are not necessary, but by interpolating the means, the cubic interpolating spline again ignores sampling variability. The result is often a curve that is extremely wavy in the regions between dose values.

The final step is to "penalize" the spline in such a way that it neither overfits (interpolation, leading to large variability in the estimated response) nor oversmooths (straight line regression,

leading to bias in the estimated response). This is done by cross-validation, a method that balances bias and variability by letting the data determine the optimal amount of smoothness. Figure 3d shows the fit of such a "cross-validated cubic smoothing spline" to these hypothetical data. The fitted curve more closely describes the underlying response than those in Figure 3b, because it is based on fewer assumptions concerning the mathematical form of the response.

The use of flexible, nonspecific, smooth functions to fit data is generally called "nonparametric regression." Note, however, that there is no clear separation between parametric and nonparametric regression methods. The difference is the degree to which the fitted curve is constrained by a mathematical model. Parametric regression is based on specific models thought to describe, or at least approximate, underlying biological mechanisms, whereas nonparametric models encompass a wider class of mathematical functions that have no such biologically based interpretation. Furthermore, nonparametric regression produces parameters; in general, the number of parameters with nonparametric regression will be larger than with a parametric model, reflecting the greater flexibility of the nonparametric function (i.e., our uncertainty about the true model).

In general, there is a trade-off involved in choosing between parametric and nonparametric regression methods. If the model is correct, we expect that the parametric model will provide more powerful inference concerning mutagenicity and potency. However, if the model is wrong, the results can be biased, possibly leading us to incorrect conclusions. There does not yet seem to be any single mathematical model for the Ames assay that a) incorporates all of the biological features of the assay, b) requires no unverifiable assumptions, and c) can be successfully

fit to the variety of results encountered in practice. The lack of such a model motivated us to investigate the possible applicability of nonparametric regression methods to Ames assay data.

Implementation

We have chosen to use the cubic smoothing spline (CSS) because it is visually smooth and is easily differentiated. The latter is important for computing potency, which requires calculation of the initial slope of the dose-response curve (see the next section for some discussion of biases in CSS estimation of potency). Detailed treatment of the CSS is available elsewhere (16,17).

The CSS is the solution $\tilde{g}(x)$ to the following penalized weighted least-squares problem: minimize

$$\frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} \frac{[Y_{ij} - g(x_i)]^2}{\sigma_i^2} + \lambda \int_{x_1}^{x_K} [g''(x)]^2 dx \quad (3)$$

over the set of twice continuously differentiable functions g with square integrable second derivative. The smoothing parameter $\lambda \geq 0$ determines where the solution lies between a straight line (as $\lambda \rightarrow \infty$) and a cubic interpolating spline (as $\lambda \rightarrow 0$). The estimator is a set of piecewise cubic polynomials

$$\tilde{g}(x) = \hat{\mu}_i + \hat{\beta}_i(x - x_i) + \hat{\gamma}_i(x - x_i)^2 + \hat{\delta}_i(x - x_i)^3 \quad (4)$$

($x_i \leq x \leq x_{i+1}$, $i = 1, \dots, K-1$), which are constrained to be continuous and to have continuous first and second derivatives. These constraints result in a reduction of Equation 4 to K parameters, which we may take to be the fitted values $\tilde{\mu}_i = \tilde{g}(x_i)$. For fixed λ , Equation 3 reduces to a set of linear equations

$$\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_K)^T = B(\lambda)\bar{y}$$

where $\bar{y} = (\bar{y}_1, \dots, \bar{y}_K)^T$ is the vector of sample means and $B(\lambda)$ is a smoothing matrix depending on the σ_i^2 and the x_i but not on \bar{y} (see the Appendix).

An obvious estimator of mutagenic potency is $\tilde{\beta}_1$, the initial slope of the piecewise curve (Eq. 4) fit to the segment between $x_1 = 0$ and x_2 , the lowest nonzero dose. Note that $\tilde{\beta}_1$ depends on the other data points as well; in fact, $\tilde{\beta}_1$ is a linear combination of $\tilde{\mu}$:

$$\tilde{\beta}_1 = a^T \hat{\mu} = a^T B(\lambda)\bar{y}$$

with a being a known vector (Appendix), and so $\tilde{\beta}_1$ is also a linear function of \bar{y} for fixed λ . The variance of $\tilde{\beta}_1$ is easily obtained using distribution theory for linear combinations of sample means.

In practice, λ and the σ_i^2 must be estimated. Because of replicates in the Ames assay, we have the usual unbiased sample variance estimates

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

which ordinarily are not available in CSS applications. The value

of λ is chosen optimally according to cross-validation by minimizing

$$CV(\lambda) = \frac{1}{n} \sum_i \frac{\sum_j (Y_{ij} - \hat{\mu}_i)^2}{s_i^2 \left(1 - \frac{b_i}{n_i}\right)^2}$$

with respect to λ , where b_i is the i th diagonal element of $B(\lambda)$ and the $\hat{\mu}_i$ depend implicitly on λ . We denote cross-validated estimators by an asterisk (*).

Although the cross-validated fitted values $\tilde{\mu}^*$ are not a linear function of \bar{y} , it may be shown that, for a fixed set of dose values, asymptotically, $\tilde{\mu}^*$ and \bar{y} have the same distribution if the number of replicate observations increases at each dose (18). This result is especially useful in providing a means of making inference concerning the initial slope. Since $\tilde{\beta}_1^*$ is a linear function of $\tilde{\mu}^*$, it is asymptotically a linear combination of \bar{y} and so has a limiting normal distribution:

$$\sqrt{n}(\tilde{\beta}_1^* - \beta_1) \xrightarrow{d} N(0, a^T \Gamma a)$$

where β_1 is the initial slope of a cubic interpolating spline fit to

the vector of true means $\mu = (\mu_1, \dots, \mu_K)^T$, $\Gamma = \text{diag} \left\{ \frac{\sigma_i^2}{v_i} \right\}$, and $0 < v_i = \lim_{n_i \rightarrow \infty} \frac{n_i}{n} < \infty$ is the limiting fraction of points at x_i . We may therefore base inference concerning the initial slope upon the standard normal z score

$$z = \frac{\tilde{\beta}_1^*}{SE(\tilde{\beta}_1^*)}$$

assuming β_1 is close to the true mutagenic potency (see the discussion of biases in the next section). Monte Carlo studies of the small sample behavior of $\tilde{\beta}_1^*$ (not shown) reveal that both its bias and variance can be strongly affected by estimation of the smoothing parameter λ . The magnitude of these effects is evaluated briefly in comparison studies discussed below.

Biases in the Potency Estimator

The validity of $\tilde{\beta}_1^*$ as a potency estimator depends on the unknown model $\mu(x)$ because $\tilde{\beta}_1^*$ estimates $\mu'(0)$ rather than $M'_M(0)$ [recall $M_n(x) = N_0 M(x)$]. In the absence of toxicity or in the case of a toxicity threshold, $\mu'(0) = M'_M(0)$. But without knowledge of $\mu(x)$, it is not possible to obtain $M_n(x)$. However, if the data are generated by Haynes-Eckardt-type models or multiple generation models, it is possible to directly estimate $M'_M(0)$ and also to calculate $\mu'(0)$, so that we may assess the impact of ignoring the survival function $S(x)$. Let the difference $bias_i = \mu'(0) - M'_M(0)$ be called the "toxicity bias." Some algebra shows that, in the case of Equation 1, $bias_i = -m_{0S}$ (3), whereas for the multigeneration Equation 2, $bias_i = -N_0 \mathfrak{S}$ where

$$\mathfrak{S} = s(1 + 3e^{-m_d/N_0} + 8e^{-2m_d/N_0} - 12e^{-3m_d/N_0})$$

The validity of $\tilde{\beta}_1^*$ also depends on how well a cubic spline approximates the true dose-response function $\mu(x)$. Let $bias_a = \beta_1 - \mu'(0)$ be called the "approximation bias"; this is the bias due to approximating the initial slope of $\mu(x)$, using the interpolating spline fit to the true means. This bias may be calculated for specific dose-response models by simulating the model and fitting the cubic interpolating spline. Figure 4 shows the relative bias $bias_a / \mu'(0)$ as a function of true mutagenic potency m_1 in Haynes-Eckardt (Eq. 1) and multigeneration (Eq. 2) models, using various values of the toxicity parameter s (0.001, upper panels; 0.01, lower panels). In these simulations, six dose groups were used with relative spacing (0, 0.015625, 0.125, 0.25, 0.5, 1); the highest dose was calculated to give either 90% or 10% survival. [The appropriate maximum dose for achieving 90% or 10% survival with the multiple generation model assuming unlimited toxicity may be calculated using probabilities derived by Margolin et al. (9).] These results show that the spline approximation bias is quite large with 10% survival, presumably because there are insufficient data for the spline to mimic the dose-response structure between the first two dose values in the low-dose portion of the curve. Increasing the quadratic term also increases the relative bias.

Finally, there will be bias due to estimation of the smoothing parameter λ ; call $bias_s = \tilde{\beta}_1^* - \beta_1$ the "smoothing bias." According to the above asymptotic distribution the smoothing bias will be small in large samples. This bias is not easily computed theoretically, but it may be ascertained from Monte Carlo studies by subtracting the toxicity and approximation biases from the total bias of the estimator, as illustrated in the next section. The total bias in $\tilde{\beta}_1^*$ as an estimate of m_1 in Equation 1 is

$$bias_{tot} = \tilde{\beta}_1^* - m_1$$

$$= [\tilde{\beta}_1^* - \beta_1] + [\beta_1 - \mu'(0)] + [\mu'(0) - m_1]$$

$$= bias_s + bias_a + bias_t$$

so $bias_s = bias_{tot} - bias_t - bias_a$. Note that $bias_t$ and $bias_a$ are functions of the unknown $\mu(x)$, so they cannot be individually ascertained except by knowledge of $\mu(x)$. On the other hand, $bias_s$ is a small sample bias that can be made small by the use of sufficiently large samples.

Comparative Studies

We now describe Monte Carlo studies designed to evaluate the CSS approach and compare the CSS to two other regression methods: nonlinear regression and the point rejection method. We also compare the fits of these methods to the two example data sets introduced earlier. The multigeneration approach of Margolin et al. (9,14) was not compared because of the difficulty of estimating the generation times.

Monte Carlo Studies

In the Monte Carlo studies we used a variety of Haynes-Eckardt and multigeneration models, with varying degrees of curvature and either 90% or 10% minimum survival, to obtain random Ames assay type data. All studies were based on five dose levels with relative spacing (0, 0.125, 0.25, 0.5, 1). Random Poisson or negative binomial variables were generated according to algorithm "NG" of Ahrens and Dieter (19). Three replicates were generated at each dose. Each experiment consisted of 2000 simulated data sets. If there was no quadratic mutagenicity term ($m_2 = 0$) in the model that generated the data, we fit a linear

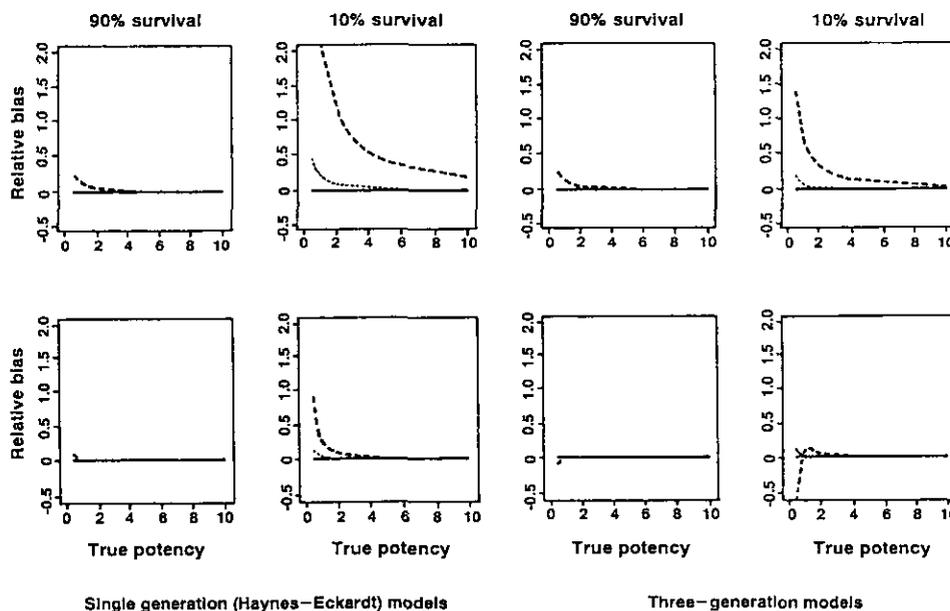


FIGURE 4. Interpolating spline relative bias in the initial slope estimator. Results are for dose responses simulated according to either Equation 1 or Equation 2, using various values of the quadratic mutagenicity term: [$m_2 = 0$ (—), $m_2 = 0.01$ (- - -), and $m_2 = 0.1$ (- · -)], survival parameter ($s = 0.001$, upper panels; $s = 0.01$, lower panels).

Table 1. Simulated comparisons among the smoothing spline, point rejection, and nonlinear regression methods: Haynes-Eckardt models with 90% survival.

Method	Parameter	$m_1 = 0, m_2 = 0.01, s = 0$		$m_1 = 2, m_2 = 0, s = 0.01$		$m_1 = 2, m_2 = 0.1, s = 0.01$	
		Distribution		Distribution		Distribution	
		P	NB	P	NB	P	NB
SS	Potency	0.18	0.18	1.65	1.59	2.12	2.07
	SSE	0.23	0.31	2.17	2.37	2.33	2.14
	CSE	0.43	0.45	2.62	4.05	3.86	4.10
	Power	0.059	0.067	0.45	0.17	0.16	0.16
PR	Potency	0.93	0.93	1.80	1.82	2.55	2.57
	SSE	0.05	0.07	0.77	0.84	0.57	0.59
	CSE	0.10	0.10	0.50	0.52	0.48	0.50
	Power	1.0	1.0	0.94	0.90	1.0	0.99
	No. of points ^a	5.0	5.0	4.8	4.8	4.9	4.9
NR	Potency	-0.11	-0.13	3.52	3.65	1.56	1.39
	SSE	0.21	0.23	1.85	2.0	2.18	2.49
	CSE	0.24	0.27	2.96	3.17	3.22	3.58
	Power	0.011	0.010	0.37	0.33	0.12	0.096
	No. of convergences ^b	1570	1632	1916	1933	1610	1545

Abbreviations: SS, smoothing spline; PR, point rejection; NR, nonlinear regression; P, Poisson; NB, negative binomial; SSE, sampling standard error; CSE, computed standard error; m_1 , linear term; m_2 , quadratic term; s , toxicity term.

^aAverage number of retained points after point rejection.

^bNumber of cases in which NR converged.

mutagenicity model with nonlinear regression [Eq. 1 with $m_2 = 0$]; otherwise we fit the full model (Eq. 1). With all three methods, the test of mutagenicity was based on asymptotic normality of the initial slope estimator divided by its standard error. Power was estimated for a one-sided level 0.05 test by the proportion of times the z score exceeded the critical value 1.645.

The first set of results (Table 1) is for data generated by Equation 1, taking $m_0 = 25$ and using various values of m_1, m_2 , and s at 90% minimum survival. Sampling standard error is the standard deviation of potency estimates from the 2000 simulated data sets in one experiment; computed standard error is the average of the 2000 individual standard errors computed from each simulated data set in an experiment. Simulated data were either Poisson or negative binomial with variance $\mu[1 + 0.005\mu]$. Because the models evaluated were nearly linear over the range of doses at which 90% minimum survival was achieved, the nonlinear regression method failed to converge about 25% of the time except in the case where $m_2 = 0$. Most notable from these studies were the positive bias and inappropriate test size of the point rejection method under the null hypothesis ($m_1 = 0$) in the presence of upward curvature, contrasted with the lack of power of the smoothing spline and nonlinear regression methods under the alternative model ($m_1 = 2$). Negative binomial errors (to mimic overdispersion) led to a reduction in power and a slight increase in standard error with each method. Results for data generated by the multiple generation model with 90% minimum survival were similar and so are not shown.

The second set of results (Table 2) is for data generated by Haynes-Eckardt and multiple generation models, using various parameter values, and with 10% minimum survival. For multiple generation models, parameters of the mutagenicity function $M(x)$ may be approximately related to parameters of the Haynes-Eckardt model $M_N(x)$ by dividing the latter by $7 \times N_0$ (13). We took $N_0 = 10^8$. All simulated models displayed

substantial curvature over the wider range of doses produced by allowing survival to decrease to 10%, and the nonlinear regression method nearly always converged. The power of all three methods was greater than in the above cases of 90% minimum survival. The nonlinear regression method performed quite well on data generated by Haynes-Eckardt models, whereas the smoothing spline and point rejection initial slope estimates were highly sensitive to the magnitude of the quadratic term.

With multiple generation models and 10% minimum survival, the smoothing spline was the only method able to closely estimate a small positive potency (last column of Table 2). The nonlinear regression method gave highly biased and variable initial slope estimates, whereas the point rejection method overestimated the initial slope when the true potency was small.

As an example of the total bias calculations for the smoothing spline (see the previous section), consider the estimate 1.65 from Table 1 ($m_1 = 2, m_2 = 0, s = 0.01$, with Poisson errors and 90% survival). The total bias is

$$bias_{tot} = \hat{\beta}_1^* - m_1 = 1.65 - 2 = -.35$$

Now

$$bias_t = -m_0 s = -25 \times .01 = -.25$$

and, from calculation of the interpolating spline fit to the true means, $bias_a = -0.01$. Thus

$$bias_s = -.35 + .25 + .01 = -.09$$

is the amount of bias in $\hat{\beta}_1^*$ due to smoothing.

Example Data Sets

Earlier, we introduced two example data sets from Ames assays of coal tar base fraction (Fig. 2a) and quinoline (Fig. 2b).

Table 2. Simulated comparisons among the smoothing spline, point rejection, and nonlinear regression methods: Haynes-Eckardt models with 10% survival.

Method	Parameter	Haynes-Eckardt, $m_1=2, s=0.01$		Three-generation, $s=0.01$		
		$m_2=0$	$m_2=0.1$	$m_1=2$		$m_1=0.1$
				$m_2=0$	$m_2=0.1$	$m_2=0.01$
SS	Potency	1.29	3.18	0.91	1.79	0.19
	SSE	0.25	0.37	0.54	0.63	0.59
	CSE	0.27	0.36	0.76	0.81	0.70
	Power	0.99	1.0	0.28	0.73	0.061
PR	Potency	1.0	4.08	0.79	1.82	0.87
	SSE	0.11	0.10	0.32	0.29	0.15
	CSE	0.11	0.11	0.26	0.22	0.16
	Power	1.0	1.0	0.63	1.0	0.99
	No. of points ^a	3.0	4.0	3.2	3.7	4.0
NR	Potency	2.01	1.97	4.35	-0.91	-1.63
	SSE	0.19	0.52	0.64	1.57	1.17
	CSE	0.18	0.51	0.93	1.97	1.48
	Power	1.0	0.97	1.0	0.14	0.001
	No. of convergences ^b	2000	2000	2000	1998	2000

Abbreviations: SS, smoothing spline; PR, point rejection; NR, nonlinear regression; P, Poisson; NB, negative binomial; SSE, sampling standard error; CSE, computed standard error; m_1 , linear term; m_2 , quadratic term; s , toxicity term.

^aAverage number of retained points after point rejection.

^bNumber of cases in which NR converged.

Table 3. Initial slope inference for example data sets.

Method	Parameter	Coal base data (8)	Quinoline data (9)
Smoothing spline	Estimate	0.52	0.18
	SE	0.13	0.50
	z Score	0.46	0.36
Point rejection	Estimate	17.0	0.24
	SE	0.8	0.06
	z Score	21.3	4.0
Nonlinear regression	Estimate	-1.0 ^a	0.20 ^b
	SE	3.5	0.07
	z Score	-0.29	2.9

^aEquation 1 with linear-quadratic mutagenicity.

^bEquation 1 with linear mutagenicity.

Table 3 displays estimated potencies obtained from the three methods compared above, along with standard errors and z scores.

As with the Monte Carlo studies, there is large bias in the point rejection slope due to an upwardly curving response in the case of the coal base data. However, the point rejection method is the most powerful in the case of the quinoline data, in which the low-dose linearity assumption appears reasonable, and all three methods gave similar estimates. The smoothing spline dose-response estimate for the quinoline data appears strange in the high-dose region because there is a large gap with no information between the highest and next highest doses (a seemingly uninteresting region of the dose response). This points out the need to consider issues of experimental design; the same total sample size applied to a narrower dose range would provide more information concerning mutagenic potency.

Conclusions

Each of the methods we compared performed best in situations where its assumptions most closely held true. The

nonlinear regression method [based on the Haynes-Eckardt model (Eq. 1)] performed well when the true model was Equation 1 and the dose range was wide enough that all of the parameters were estimable from the data. The point rejection method was quite powerful when the underlying dose response was nearly linear, but otherwise it overestimated mutagenic effects. The smoothing spline performed best in situations when the nonlinear regression was inappropriate because of the wrong model (e.g., with multigeneration models) and when the point rejection method was wrong due to curvature in the initial dose response (e.g., with nonlinear models and wide dose ranges). No single method performed well under all situations.

The smoothing spline approach was designed to provide an alternative to model-based regression methods when the underlying model is nonlinear and does not follow Equation 1. We observed that, when the dose range is selected so that minimum survival is at least 90%, even multigeneration models are nearly linear over the entire dose range, so that the point rejection method produces quite reasonable results. In this sense, the smoothing spline is rather disappointing because it appears

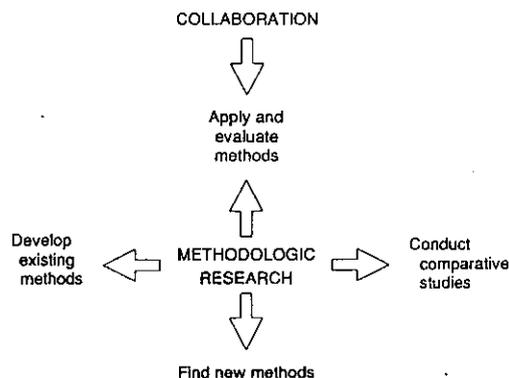


FIGURE 5. Illustration of needs for further research into methodologies for analyzing Ames assay data.

that we may achieve model robustness using the point rejection method so long as adequate dosing experiments are performed first. However, a greater understanding of the Ames assay on the part of statisticians is needed before it can be said whether this will be true generally in practice. For example, day-to-day variation in Ames assay results decreases the effectiveness of dosing experiments (B. H. Margolin, personal communication); this may lead to difficulties in attempts to routinely achieve approximate linearity.

Based on the lack of a single acceptable method, it appears that more work is needed to further develop existing methods or to find new methods that will perform well with Ames assay data without being so closely tied to model assumptions. In addition, further comparisons among methods are needed to understand in which circumstances the various methods are appropriate. Such studies, as well as comparisons on a number of actual Ames data sets, should be performed in conjunction with laboratory investigators so that situations that are most relevant in practice may be identified and studied. Figure 5 summarizes the need for further research into statistical methods suitable for use with the Ames assay. It emphasizes the important role that collaboration and application should play in this work. Statisticians can also contribute by advising researchers on how they can design better experiments so as to obtain the sort of data needed for robust analyses. Finally, it would be useful to have methods for comparing goodness of fit among the variety of models discussed in this paper. It is hoped that more statisticians will become interested in these problems and that greater collaboration between statisticians and laboratory investigators will occur in the future.

Appendix

First, let λ be fixed. The solution to Equation 3 is a piecewise cubic polynomial (16), so Equation 3 may be rewritten: minimize

$$\frac{1}{n}(Y - J\mu)^T S^{-1}(Y - J\mu) + \lambda\mu^T \Omega \mu$$

where J is the expansion matrix

$$J = \begin{pmatrix} 1_{n_1} & 0 & \dots & 0 \\ 0 & 1_{n_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1_{n_k} \end{pmatrix}$$

(20), $Y = (Y_{1,1}, \dots, Y_{1,n_1}; Y_{2,1}, \dots, Y_{2,n_2}; \dots; Y_{k,1}, \dots, Y_{k,n_k})^T$ is the vector of all observations, Ω is the "penalty matrix," which depends only on the x_i , and $S = \text{diag}\{\text{Var}(Y_{ij})\}$ (i.e., a diagonal matrix of the s_i^2 suitably expanded). Differentiating and setting the result equal to zero, we obtain

$$\begin{aligned} \mu &= (J^T S^{-1} J + n\lambda\Omega)^{-1} J^T S^{-1} Y \\ &= (S_y^{-1} + n\lambda\Omega)^{-1} S_y^{-1} \bar{y} \\ &= B(\lambda)\bar{y} \end{aligned}$$

with $S_y = \text{diag}\left\{\frac{s_i^2}{n_i}\right\}$. The $\bar{\mu}$ may be computed directly or via

the equivalent algorithm "SMOOTH" of deBoor (16) that solves for the third-order coefficients γ_i in Equation 4.

For estimating λ by cross-validation, Silverman (21) has recommended golden section search. At each iteration, we compute $CV(\lambda)$ using $\bar{\mu}$ based on the current value of λ in the search. Occasionally, because of unequal spacing between dose points, there is a tendency to interpolate nearby means. This produces wavy dose-response curve estimates that are unlikely to be acceptable to investigators. In such cases, we find that minimizing the weighted cross-validation score

$$CV_w(\lambda) = \sum_i \frac{w_i \sum_j (Y_{ij} - \bar{\mu}_i)^2}{n_i s_i^2 \left(1 - \frac{b_i}{n_i}\right)^2}$$

using the interval-length weights (D. Ragozin, personal communication)

$$\begin{aligned} w_1 &= \frac{(x_2 - x_1)}{2} \\ w_i &= \frac{x_{i+1} - x_{i-1}}{2}, \quad i=2, \dots, K-1 \\ w_K &= \frac{(x_K - x_{K-1})}{2} \end{aligned}$$

often produces satisfactorily smooth dose-response estimates.

To compute the initial slope estimator, we use the fact that Ω may be written as $\Omega = QR^{-1}Q^T$, where Q and R are tridiagonal matrixes (16). Then, using linear relationships among the coefficients of the piecewise polynomial (Eq. 4) (16,22),

$$\beta_1 = \frac{\bar{\mu}_2 - \bar{\mu}_1}{x_2} - \frac{x_2}{6} \omega^T \bar{\mu} = a^T \bar{\mu}$$

where ω is the first column of QR^{-1} and

$$a = \begin{pmatrix} -\frac{1}{x_2} \\ \frac{1}{x_2} \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \frac{x_2}{6} \omega$$

This work was supported in part by U.S. Public Health Service (U.S. PHS) training grant no. T-32-CA-09168 while the first author was a graduate student in the Department of Biostatistics at the University of Washington, and in part by U.S. PHS grant CA-40644 to N.E.B. We gratefully acknowledge Elaine Faustman for invaluable advice, Leslie Bernstein, who kindly provided the point rejection program, James Trosko for a critical reading of an earlier draft, and Ritsko Sato, who helped with the preparation of figures and with numerous administrative tasks. Finally, we thank David Hoel and Akira Sakuma for the invitation to present this work.

REFERENCES

1. Ames, B. N., McCann, J., and Yamasaki, E. Methods for detecting carcinogens and mutagens with the *Salmonella*/mammalian microsome mutagenicity test. *Mutat. Res.* 31: 347-364 (1975).
2. Maron, D., and Ames, B. M. Revised methods for the *Salmonella* mutagenicity test. *Mutat. Res.* 113: 173-215 (1983).
3. Breslow, N., and Kaldor, J. Statistical analysis of data from *in vitro* assays of mutagenesis. In: *Long-Term and Short-Term Assays for Carcinogens: A Critical Appraisal* (R. Montesano, H. Bartsch, H. Vainio, J. Wilbourn, and H. Yamasaki, Eds.), IARC Scientific Publications No. 83, International Agency for Research on Cancer, Lyon, 1986, pp. 457-481.
4. Chu, K. C., Patel, K. M., Lin, A. H., Tarone, R. E., Linhart, M. S., and Dunkel, V. C. Evaluating statistical analysis and reproducibility of microbial mutagenicity assays. *Mutat. Res.* 85: 119-132 (1981).
5. Simpson, D. G., and Margolin, B. H. Recursive nonparametric testing for dose-response relationships subject to downturns at high doses. *Biometrika* 73: 589-596 (1986).
6. Horn, L., Kaldor, J., and McCann, J. A comparison of alternative measures of mutagenic potency in the *Salmonella* (Ames) test. *Mutat. Res.* 109: 131-141 (1983).
7. Margolin, B. H. Potency measures in mutagenicity research. *Environ. Health Perspect.* 101: 00-00 (1992).
8. Myers, L. E., Sexton, N. H., Southerland, L. I., and Wolff, T. J. Regression analysis of Ames test data. *Environ. Mutagen.* 3: 575-586 (1981).
9. Margolin, B. H., Kaplan, N., and Zeiger, E. Statistical analysis of the Ames *Salmonella*/microsome test. *Proc. Natl. Acad. Sci. U.S.A.* 78: 3779-3783 (1981).
10. Haynes, R. H., and Eckardt, F. Analysis of dose-response patterns in mutation research. *Can. J. Genet. Cytol.* 21: 277-302 (1979).
11. Stead, A. G., Hasselblad, V., Creason, J. P., and Claxton, L. Modeling the Ames test. *Mutat. Res.* 85: 13-27 (1981).
12. Breslow, N. E. Extra-Poisson variation in log-linear models. *Appl. Stat.* 33: 38-44.
13. Bernstein, L. J., Kaldor, J., McCann, J., and Pike, M. C. An empirical approach to the statistical analysis of mutagenesis data from the *Salmonella* test. *Mutat. Res.* 97: 267-281 (1982).
14. Margolin, B. H., Kim, B. S., and Risko, K. J. The Ames *Salmonella*/microsome mutagenicity assay: issues of inference and validation. *J. Am. Stat. Assoc.* 84: 651-661 (1989).
15. Weinstein, I. B. Mitogenesis is only one factor in carcinogenesis. *Science* 251: 387-388 (1991).
16. deBoor, C. A *Practical Guide to Splines*. Springer-Verlag, New York, 1978.
17. Eubank, R. L. *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, New York, 1988.
18. Cologne, J. B. *Nonparametric regression analysis of the Ames mutagenicity assay*. Ph.D. dissertation, Department of Biostatistics, University of Washington, Seattle, WA, 1989.
19. Ahrens, J. H., and Dieter, U. Computer methods for gamma, beta, Poisson, and binomial distributions. *Computing* 12: 223-246 (1974).
20. Ansley, C. F., and Kohn, R. Spline smoothing with repeated values. *J. Stat. Comput. Simul.* 25: 251-258 (1986).
21. Silverman, B. W. A fast and efficient cross-validation method for smoothing parameter choice in spline regression. *J. Am. Stat. Assoc.* 79: 584-589 (1984).
22. Wegman, E. J., and Wright, I. W. Splines in statistics. *J. Am. Stat. Assoc.* 78: 351-365 (1983).