Journal of Coastal Research	25	2	366–379	West Palm Beach, Florida	March 2009

Toward Parsimony in Shoreline Change Prediction (I): Basis Function Methods

L. Neil Frazer, Ayesha S. Genz, and Charles H. Fletcher

Department of Geology and Geophysics
School of Ocean and Earth Science and Technology
University of Hawaii
1680 East West Road
Honolulu, HI 96822 U.S.A.
neil@hawaii.edu

ABSTRACT

FRAZER, L.N.; GENZ, A.S., and FLETCHER, C.H., 2009. Toward parsimony in shoreline change prediction (I): basis function methods. *Journal of Coastal Research,* 25(2), 366–379. West Palm Beach (Florida), ISSN 0749-0208.

Single-transect methods of shoreline change prediction are unparsimonious, *i.e.,* they tend to overfit data by using more parameters than necessary because they assume that both signal and noise at adjacent transects are independent. Here we introduce some new methods that reduce overfitting by expressing change rate as a linear sum of basis functions. In the method of IC-binning, the basis functions are boxcars—an information criterion is used to assign contiguous alongshore locations into bins within which change rate is constant; the resulting rate is discontinuous but may be useful for beach management. In the polynomial method, the basis functions are polynomials in alongshore distance, and the change rate varies continuously along the beach. In the eigenbeaches method, the basis functions are the principal components of the matrix of shorelines. To choose the number of basis functions in each method, and to compare methods with each other, we use an information criterion. We apply these new methods to shoreline change on Maui Island, Hawaii, briefly here, and in more detail in a companion paper. The polynomial method works best for short beaches with rates that vary slowly in the alongshore direction while eigenbeaches works best for shorelines that are long, or have rates that vary rapidly in the alongshore direction. The Schwarz information criterion and the AICu version of the Akaike information criterion performed well in tests on real data and noisy synthetic data.

ADDITIONAL INDEX WORDS: *Coastal erosion, shoreline change rates, Hawaii beaches, information criterion, AICc, AICu, AICm, SIC, gMDL.*

INTRODUCTION

Storms, sea level rise, human impacts to sediment availability, and other causes of shoreline change result in extensive damage to global coastlines. The widespread nature of this problem means that coastal managers need good tools and scientific databases to develop management policies. Coastal managers in many states are implementing quantitative methods, such as shoreline change rates, to identify setbacks for construction in erosion hazard zones (National Academy of Sciences, 1990).

The modeling of shoreline change has evolved through a succession of research efforts. Coastal scientists and managers once preferred the end-point rate method, which uses only two shoreline positions to calculate rate of change at each alongshore location. Most coastal scientists now favor least squares regression over end-point rate because least squares utilizes data from all shoreline positions to calculate shoreline change (*e.g.,* Crowell, Douglas, and Leatherman, 1997; Galgano and Douglas, 2000; Honeycutt, Crowell, and Douglas, 2001). Recently, Genz *et al.* (2007) incorporated measurement and positional uncertainties into the shoreline change rate procedure; they thoroughly tested various forms of weighted least squares, least absolute deviation, and least median of squares on real and synthetic data with Gaussian and non-Gaussian errors; they also introduced a binning method that recognizes the lack of independence of data from adjacent transects. The present paper and its companion build on that effort.

Shoreline data sets seldom have as much data as the user would like. Typically, change rates are calculated at shore-normal transects that contain between 5 and 10 historical shoreline positions unevenly spaced through time. These data sets often have large scatter due to short-term beach changes from storms, as well as seasonal and tidal effects, and this scatter can mask the long-term change rate signal, resulting in misleading predictions of hazard zones (Galgano and Douglas, 2000; Honeycutt Crowell, and Douglas, 2001; Zhang, Douglas, and Leatherman, 2002).

In this situation, it is important to be certain that data are not being overfitted. The classical example of overfitting is the use of an $n - 1$ degree polynomial to fit n noisy data points: The fit is perfect, but if the fitted polynomial is used for extrapolation or interpolation, the results can be nonsense. Similarly in shoreline change, if the rate varies rapidly in the alongshore direction, and that rapidly varying rate is

DOI: 10.2112/06-0756.1 received 14 September 2006; accepted in revision 28 February 2008.

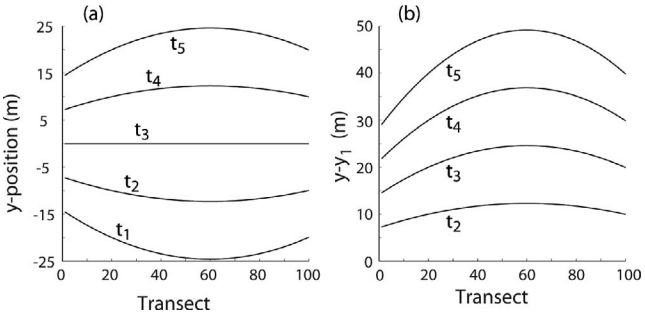

Figure 1. (a) A beach at successive times, as determined by 100 transects. (b) The first beach has been subtracted from subsequent beaches, revealing the rate at each transect. If each transect is regarded as independent, the beach is a vector in a space of dimension 100. However, the rate is well modeled by $r = a + bx + cx^2$, so this beach is better regarded as a vector in the three-dimensional space of [*a, b, c*].

used for long-term prediction, the predicted shoreline may have more variability than the smoothing processes of nature are likely to allow. We expect a future shoreline to vary in the alongshore direction, just as present shorelines do, but the high spatial-frequency part of that variability is less predictable than the low spatial-frequency part.

The importance of avoiding overfitting is sometimes referred to as Occam's razor, or the parsimony principle—a model with more parameters than are strictly necessary to fit the data is referred to as an unparsimonious model. Fenster, Dolan, and Elder (1993) were the first to apply a modern information criterion to ensure parsimony in shoreline change modeling. We extend their pioneering efforts by including the alongshore dimension of shoreline change in the model basis functions by incorporating the dependence of adjacent transects in both model and noise, and by testing a variety of modern information criteria.

Figure 1 illustrates the parsimony principle for an idealized beach determined by 100 transects surveyed at 5 different times. If each transect is regarded as independent, the rate is a point in a 100-dimensional space. However, this particular beach satisfies the relation

$$y(x,t) = y(x,t_1) + (t - t_1)(a + bx + cx^2),$$

so its rate is more parsimoniously regarded as a point in a three-dimensional space [*a, b, c*]. The functions 1, *x,* and x^2 are said to be the *basis functions* for this beach, and the parameters *a, b,* and *c* are the *coefficients* of the basis functions. In practical work, simple powers of *x* are inefficient basis functions, and orthogonal polynomials are preferred; however, the principle is the same. The example of Figure 1 suggests that the model selection problem consists of two parts: first one must decide what class of basis functions to use, and then how many basis functions are required by the data. If the basis functions are well chosen, few of them will be required, and the description of the shoreline change is parsimonious. However, it is not objective procedure to model a shoreline using one class of basis functions and then select linear combinations of the original basis functions as new basis functions. Basis functions must have a natural mathematical ordering, although, as we shall see, that order can be determined by the data.

In the following discussion, we apply the parsimony principle to the shoreline change prediction problem using objective measures of parsimony, called information criteria (IC), to compare various models. An IC is a statistic that tells when data are being fitted to an overly complex model. An IC is used the way an adjusted R^2 statistic has traditionally been used; the main difference is that an IC can be used to compare models that are not necessarily nested; also, the ICs we use perform much better than the adjusted R^2 (McQuarrie and Tsai, 1998). The value of any IC decreases with goodness of fit and increases with model complexity; the model with the lowest IC score is thus an optimal compromise between parsimony and goodness of fit, among the models considered.

Our first application is to the single-transect and binning methods offered by Genz *et al.* (2007). Their binning method is here referred to as t-binning, because Student's *t*-distribution is used to find the bins. We offer an improvement, called IC-binning, in which an IC is used to find the bins and to decide how many bins are justified by the data. Next, we introduce the polynomial method, in which prediction is carried out by fitting polynomials to the data, as in the example of Figure 1. We use Legendre polynomials (suitable for shorelines with ends) and trigonometric functions (suitable for an island shoreline with no endpoints), but other types of polynomials could be used.

Finally, we introduce the eigenbeaches method, in which the basis functions used to fit the data are obtained from the data themselves. In principle, all the methods considered here can be used to model acceleration (including deceleration) of shoreline change rate. However, acceleration is difficult to detect in most data sets and is seldom meaningful unless the model has relatively few parameters, so we consider acceleration here only in the polynomial and eigenbeaches methods.

All our methods use generalized least squares to find parameter values. Generalized least squares (GLS) is the extension of ordinary least squares necessary to accommodate correlated data errors. The shoreline surveys analyzed in this paper were done years apart, so we assume that errors at different times are uncorrelated. However, the beach locations measured in each survey are very close in the alongshore direction, so we allow errors to be spatially correlated as suggested by the data. Like ordinary least squares, GLS assumes that the noise process is Gaussian, so it is not robust to erroneous leverage points (Rousseuw, 1984; Rousseuw and Leroy, 1987). Robustness in the context of the single-transect method was carefully examined by Genz *et al.* (2007) using real and synthetic data, and their results can be incorporated into our methods in a straightforward way, although computation times will then be much longer; a later section of the paper explains how this can be done. Our focus is on parsimony and models for shoreline change that allow spatially correlated rates and noise.

For each method, we use GLS to find the best values for a parameter vector of given length, and then use an IC to decide what length parameter vector is appropriate. Finally, an

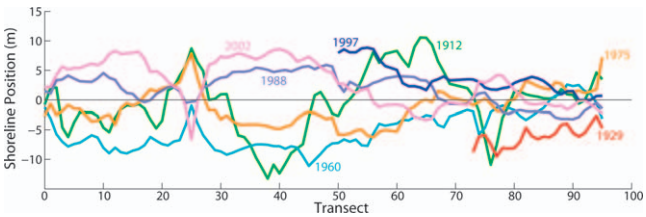

Figure 2. Shoreline data from Waihee, Maui, with baseline removed.

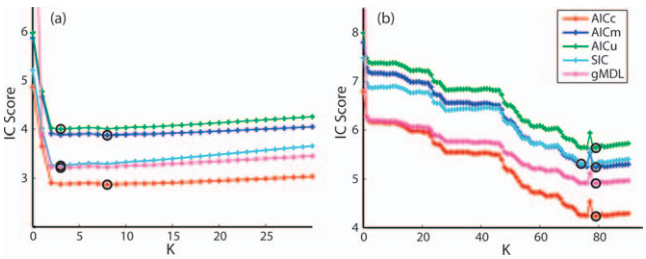

Figure 3. (a) Information criteria values for LX applied to synthetic data containing three basis functions plus noise; three out of five ICs identified the correct model, which was strongly identifiable. (b) Information criteria values for LX applied to real data from Waihee, Maui. The spike in rate near transect 25 makes the LX model only weakly identifiable.

IC is used to decide which method fits a particular beach best without overfitting the data. We assume that the method that fits the data best without overfitting, and satisfies cross-validation tests (prediction of known shorelines), is the best method to predict a future shoreline position. Several information criteria have been extensively used for scientific modeling; because they do not always agree, we test five of them.

Our data example, Waihee (pronounced why-hay-eh) beach on the island Maui, shown in Figure 2, was chosen because it illustrates nicely how several of our methods fail when used inappropriately. Waihee is relatively long, compared with other Maui beaches, and it includes a rapidly accreting point, near transect 25, so that in the alongshore direction change rate behaves somewhat like a delta function.

INFORMATION CRITERIA

An information criterion (IC) is a test statistic, or score, that determines the best model from a given set of models that are not necessarily nested: the lower the score, the better the model. The corrected Akaike information criterion (AICc) (Hurvich and Tsai, 1989; Sugiura, 1978) is the most widely used IC, and we tested it extensively. We also tested other, similar statistics such as the Schwarz information criterion (SIC) (Schwarz, 1978), and the minimum description length statistic (MDL) (Rissanen, 1978). Our MDL is the mixture form analyzed by Hansen and Yu (2001, 2003), and referred to by them as gMDL. Because most of the discussion applies directly to these other statistics with little change, we treat AICc in some detail and present other methods mainly by formula so that their differences from AICc are clear. In statements that apply to any information criterion, we use the acronym IC. We use the IC score in two ways: to find the optimal model within a class of models and to compare the best models between classes. For example, in the polynomial methods we use IC scores to find the polynomial that best fits the data, and in the binning methods we use IC scores to find the bin configuration that best fits the data. Then we use IC scores to compare the best binning model with the best polynomial model.

In comparing ICs it is important to note that for any IC the quantity $a + b \times \text{IC}$ has the same minimum as IC, where a and b are any two positive numbers; thus two IC formulas that look very different can represent the same IC, and the literature is not standard. Our formula for AICc, derived heuristically in Appendix A, is an extension of the formula in Burnham and Anderson (2002, p. 66; divide their formula by N) to data with correlated noise. Let d be a column vector of data containing estimated shorelines at various times, and let $G\hat{m}$ be the data vector predicted from a particular model of shoreline change with system matrix G and best fit parameter vector $\hat{m}$. Then the AICc score of the model is given by

$$\text{AICc} = \log(\text{RSS}/N) + 2K/(N - K - 1) \quad (1a)$$

$$\text{RSS} = |\tilde{C}_{dd}|^{1/N}(d - G\hat{m})^T\tilde{C}_{dd}^{-1}(d - G\hat{m}). \quad (1b)$$

In these equations, RSS is effectively a sum of squared residuals; N is the number of data points (length of d); the best-fit parameter vector $\hat{m}$ has length M; K is the number of parameters in the statistical model (if there is only one variance parameter, $K = M + 1$); and $\tilde{C}_{dd}$ is an $N \times N$ estimated data covariance matrix with determinant $|\tilde{C}_{dd}|$. The determinant property $|aA| = a^N|A|$ shows that RSS is invariant to scaling of $\tilde{C}_{dd}$.

The main thing to notice about the AICc formula is that as the model improves, the first term declines with the residuals, while the second term is a penalty that increases with the number of parameters in the model. The best model is the one that lowers the residuals the most for the least penalty. The argument of the logarithm in an IC formula is not required to be dimensionless because a change of units from meters to centimeters, for example, adds a constant log(10,000) to the IC formula without changing the location of the minimum.

Our formulas for AICm, AICu, and SIC differ from our AICc formula only in the penalty terms. Our AICm is the AICc_k of McQuarrie and Tsai [1998, p. 22, their equation (2.14)], and our formula for AICu is the AICu of McQuarrie and Tsai [1998, p. 32, their equation (2.18)]. Our formula for SIC (Schwarz, 1978) is standard in the literature. The formulas are:

$$\text{AICm} = \log(\text{RSS}/N) + (N + K)/(N - K - 2),$$

$$\text{AICu} = \log(\text{RSS}/N) + \log[N/(N - K)] + (N + K)/(N - K - 2),$$

$$\text{SIC} = \log(\text{RSS}/N) + K\log(N)/N.$$

The formula for gMDL is more complex and is given in Appendix B. Figure 3 shows the values of different ICs *vs.*

Table 1. *ΔIC scores for best fit models of different classes applied to the data from Waihee, Maui, shown in Figure 1. Four ICs favor the EXT model, with EX a close second, and the remaining IC favors EX, with EXT a close second.*

Method	ΔAICc	ΔAICm	ΔAICu	ΔSIC	ΔgMDL
Single-transect	2.700	2.705	2.969	3.185	1.465
T-BIN	1.317	1.319	1.370	1.427	1.039
IC-BIN	0.724	0.725	0.767	0.809	0.232
LX	4.797	4.801	5.031	7.123	6.662
RX	4.762	4.766	5.010	5.225	5.284
EX	0.118	0.118	0.115	0.107	0
LXT	3.965	3.976	4.457	3.439	4.557
RXT	3.738	3.766	4.572	3.342	4.995
EXT	0	0	0	0	0.170

number of parameters for the LX method (see the following discussion) applied to shoreline data from Waihee, Maui. In comparing models with an IC, only the difference in IC values is meaningful. Accordingly, the usual practice is to subtract the IC of the best (lowest IC) model from the scores of each model to get the △IC value for each model. The best model always has a △IC of zero. Table 1 gives the △IC scores for various methods and ICs applied to data for Waihee, Maui.

It can be seen that as more parameters are added (increasing K), the penalty terms of each IC increase at different rates. For a fixed number of model parameters, as the number of data points N approaches infinity, the penalty term of AICc goes to zero rapidly, the penalty term of SIC goes to zero more slowly, and the penalty terms of AICm and AICu go to unity. Therefore, for very large amounts of data, AICu and AICm are most parsimonious, followed by SIC and AICc. Another clue to behavior is obtained by setting $K = N - 3$; then AICu again has the largest penalty function, followed by AICm, AICc, and SIC.

Parsimony is essential in an IC, but there are additional desiderata. One desideratum is *consistency*: the property that as the number of data goes to infinity, the true model will be picked with probability 1 if it is among the models being considered. Another desideratum is *asymptotic efficiency* (Shibata, 1980), the property that as the number of data goes to infinity, the model nearest to the true model (nearest in the sense of least squares or of Kullback-Leibler discrepancy) will be picked with probability 1. Consistency is a desirable feature for problems in which there is a well-verified theory so that the basis functions are known; many problems in chemistry and physics are in this category. Asymptotic efficiency is desirable when the dimension of the true model is likely to be large, and the basis functions are poorly known. SIC and gMDL are consistent; AICc and AICm are asymptotically efficient; and AICu steers a middle course between consistency and asymptotic efficiency (McQuarrie and Tsai, 1998, p. 43).

Table 2. *Degrees of freedom for all methods. I = no. of transects, J = no. of time points at each transect, T_b = no. of transects in a bin, N_b = no. of bins, K_1 and K_2 = no. of terms.*

Method	Degrees of Freedom, ν	IC Parameter Count, K
Single-transect	$J - 2$	$3I$
Binning	$T_b(J - 1) - 1$	$I + 2N_b + 1$
PXT	$N - (I + K_1 + 2K_2 + 1)$	$I + K_1 + 2K_2 + 2$
Eigenbeaches	$N - (I + K_1 + 2K_2 + 1)$	$I + K_1 + 2K_2 + 2$

Another important concept is model *identifiability*. A mark of a strongly identifiable model is that few basis functions are needed. An example of a weakly identifiable model is one whose coefficients decline slowly with increasing order. Identifiability can be examined for particular data sets by plotting the coefficients of unit-energy basis functions. For example, if the coefficients decline abruptly at a certain order, the model is strongly identifiable. It can be seen that a poor choice of basis functions will make a strongly identifiable model appear to be weakly identifiable. For example, a beach with an abrupt change of rate in the middle will be strongly identifiable with EX and IC-binning but only weakly identifiable with LX (methods given later in this article). On the other hand, a beach along which rate varies smoothly will be strongly identifiable with LX and EX but weakly identifiable with IC-binning. If the basis functions are correctly chosen, and the model is strongly identifiable, the IC should have a well-defined minimum at the correct combination of basis functions.

McQuarrie and Tsai (1998) tested a variety of ICs (though not gMDL, which is new) in least squares regression for a large suite of randomly generated system matrices, with both strongly and weakly identifiable models. They emphasize that an IC is a statistic, like mean or variance, and that with any data set there is a finite probability that a given IC will pick the wrong model. They calculated the probability that an IC will underfit or overfit for various values of N and K. Our tests on real and synthetic data are consistent with the conclusions of McQuarrie and Tsai (1998, table 2.1) for strongly identifiable models. To quote a specific example, for $K = 6$ and $N < 100$, AICu has the lowest probability of overfitting, but it has a tendency to underfit as N reaches 100. SIC is less likely to underfit as N approaches 100, but it is more likely to overfit for smaller values of N. In the context of shoreline change, if data are to be modeled independently at each transect, AICu is the preferred IC; for the large-N models of this paper (IC-binning LX/T and EX/T), SIC performed about as well as AICu.

Counting the number of parameters for purposes of IC can be confusing when models are not nested. Table 2 summarizes the count for various methods. For the single-transect method (see the following discussion), the number of parameters is three times the number of transects because three quantities are being estimated at each transect: the change rate, an intercept (implicit because of a baseline shift), and

the variance. For binning methods, the number of parameters is twice the number of bins (because there is a rate and variance in each bin), plus the number of transects (because there is an intercept for each transect), plus one for the correlation length of the spatial noise process. For the PXT and eigenbeaches methods, the number of parameters is the number of transects (because there is one intercept for each transect), plus the number of basis functions plus one for the single variance scaling factor, plus one for the correlation length of the alongshore noise. In practice, the number of parameters is invariably greatest for the single-transect method and least for the eigenbeaches method.

We noted earlier that when any IC is replaced by $a + b \times$ IC for positive a and b, it will select the same model because the expression $a + b \times$ IC takes its minimum at the minimum of IC. However, there is an application in which the multiplier b is important: Suppose the ith model has IC value IC_i; then the quantity $p_i \propto \exp(-IC_i/2)$ can be regarded as the relative probability of model i. This only works if the IC is scaled so that the first term (the RSS term) is -2 times the log(likelihood), which is how that term originated (see Appendix A).

BASELINE AND TIME ORIGIN

Beach data consist of beach locations y_{ij} at alongshore locations x_i and times t_j; here location means the distance of the shoreline from a given baseline. The transect index i runs from 1 to I, and the time index j runs from 1 to J; in other words, $y_{ij} = y(x_i, t_j)$. Because older shoreline surveys are thought to be less reliable than modern surveys, we assume here that data variance depends on time, but not on x, so the covariance matrix has a block for each transect.

To condition the matrices in our numerical procedures, it is helpful to shift observations to a new baseline and time origin generated from the data themselves. The new baseline is a weighted average of shorelines, with weights from the relative survey errors $\tilde{\sigma}_j$. The new baseline and time origin, are given, respectively, by

$$\bar{y}_i = \left(\sum_j \tilde{\sigma}_j^{-2} y_{ij}\right)\left(\sum_j \tilde{\sigma}_j^{-2}\right)^{-1} \quad \text{and}$$

$$\bar{t} = \left(\sum_j \tilde{\sigma}_j^{-2} t_j\right)\left(\sum_j \tilde{\sigma}_j^{-2}\right)^{-1}.$$

The relative survey errors $\tilde{\sigma}_j$ are calculated by the method in Fletcher *et al.* (2003) and Genz *et al.* (2007) from seven types of data error; in our calculations they are *a priori* estimates and are therefore not included in the parameter counts for an information criterion.

As shown in Figure 4, if the model is linear with time, the baseline shifts result in a fit with zero intercept. If the model includes a t^2 time variation, conditioning of the system matrix is improved by the change of baseline and time origin, but an intercept must now be included in the model because t^2 and 1 are not linearly independent functions. Inclusion of a t^2 dependency in a model is necessary to address the question of whether change rates are increasing or decreasing, and for that purpose a good model is

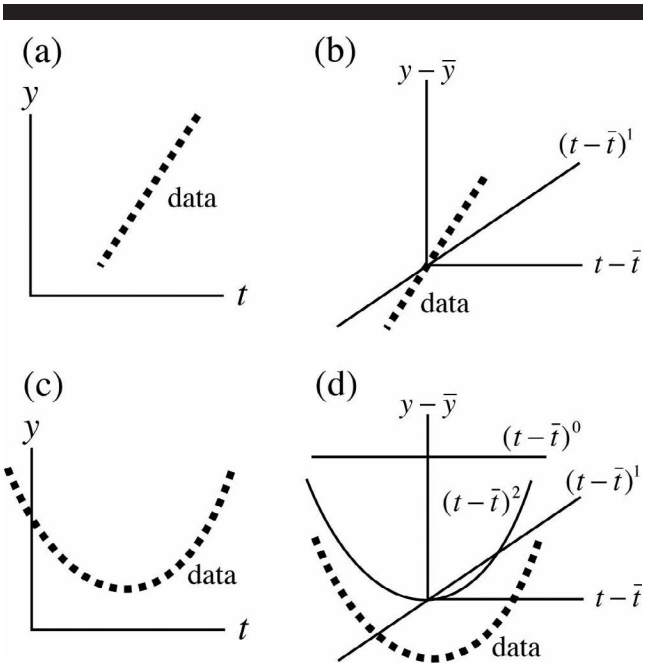

Figure 4. Our baseline and time-origin shift linear data (a) so that the best-fit WLS line has zero intercept, as in (b). However, quadratic data (c) are not shifted to zero-intercept (d), so the shifted data cannot be modeled solely by the first and second powers of $t - \bar{t}$. Models with acceleration terms therefore also require an equal number of intercept terms. With acceleration in the model, the baseline shift and time shift are still useful because they condition the matrix whose inverse is calculated in the least squares problem.

$$y_j - \bar{y} = m^{(0)} + (t_j - \bar{t})m^{(1)} + \frac{1}{2}(t_j - \bar{t})^2 m^{(2)}.$$

Note that acceleration is constant in this model, and that estimates of acceleration and rate are centered at $t = \bar{t}$. For the single-transect method, we do not look for acceleration because it is usually excluded by the IC, but we allow acceleration in the PXT and eigenbeaches methods. If a model with acceleration terms has a lower IC score than any model without such terms, that is evidence of acceleration. Note that although the time functions $t - \bar{t}$ and $(t - \bar{t})^2/2$ are orthogonal in the usual inner product on any interval symmetric about the origin, their orthogonality is not exact in the case of sampled, weighted data. Accordingly, models with and without acceleration will have slightly different rates at $\bar{t}$.

GENERALIZED LEAST SQUARES

To reduce repetition, we collect here the generalized least squares (GLS) relations that will be used to fit data and to predict 50-year shoreline positions for various methods. Our process model is

$$d = GM + n, \tag{2}$$

in which d is an $N \times 1$ data vector of beach data, m is an $M \times 1$ parameter vector, G is an $N \times M$ system matrix containing the basis functions, and n is an $N \times 1$ column vector of samples from a zero-mean, Gaussian noise process with

covariance matrix C_{dd}. This GLS problem can be transformed to an OLS problem by left-multiplying Equation (2) by $C_{dd}^{-1/2}$, but we work with untransformed data, so the covariance matrix is explicit in formulas. Shoreline data in the data vector are arranged in different ways for different methods, but if the data vector consists of time blocks, the covariance matrix is block diagonal. Because the covariance is unknowable—to estimate it directly, we would have to repeat the same experiment many times—we begin with an *a priori* estimate $\tilde{C}_{dd}$, then use the data to adjust the estimate. The estimated parameter vector is

$$\hat{m} = (G^T \tilde{C}_{dd}^{-1} G)^{-1} G^T \tilde{C}_{dd}^{-1} d, \tag{3}$$

and the best estimate proportional to $\tilde{C}_{dd}$ is

$$\hat{C}_{dd} = \hat{\alpha}_\nu \tilde{C}_{dd}, \tag{4}$$

in which the constant of proportionality is

$$\hat{\alpha}_\nu = (N - M)^{-1} (\hat{d} - G\hat{m})^T \tilde{C}_{dd}^{-1} (\hat{d} - Gm). \tag{5}$$

Here the subscript ν is a shorthand for the degrees of freedom, $N - M$, and a reminder to ensure that the division by $N - M$ is not forgotten during calculations. The parameter covariance matrix, often referred to as the model covariance matrix, is obtained from the estimated data covariance matrix by the relation (*e.g.,* Menke, 1989)

$$\hat{C}_{mm} = (G^T \hat{C}_{dd}^{-1} G)^{-1}. \tag{6}$$

Having estimated the parameters of the shoreline model, we extrapolate to predict the shoreline location at some future time t_f. A future shoreline location at alongshore coordinate x_i is just a function $y_i = q_i^T m$ where $q_i = q_i(t_f)$ is a column vector of length M, which we refer to as a prediction kernel. The future shoreline location thus has estimator

$$\hat{y}_i = y_i + q_i^T \hat{m} \tag{7a}$$

with estimated variance given by

$$\hat{\sigma}_i^2 = q_i^T \hat{C}_{mm} q_i. \tag{7b}$$

A $100(1 - \varepsilon/2)\%$ confidence interval for y is given by

$$\hat{y}_i = \bar{y}_i + q_i^T \hat{m} \pm t(\nu, 1 - \varepsilon/2) \hat{\sigma}_i \tag{7c}$$

in which $t(\nu, 1 - \varepsilon/2)$ is the value of Student's t-distribution with $\nu = N - M$ degrees of freedom (Draper and Smith, 1998). Table 2 gives the degrees of freedom and IC parameter counts for the different methods. Occasionally it is desirable to estimate change rate at a particular time, and this is done via relations analogous to Equations (7a), (7b), and (7c), using prediction kernels given later in the paper.

We estimate spatial covariance from the data themselves via the following procedure, which assumes that sources of error do not vary in the alongshore dimension, and that they vary with time only in magnitude, not scale length. First, we fit the data using the single-transect method (see following discussion) with the diagonal matrix $\tilde{C}_{dd} = \text{diag}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \hat{\sigma}_J^2)$ (*i.e.,* weighted least squares). Then we estimate the spatial correlation distance L by fitting a decaying exponential to the autocorrelation of the data residuals computed according to the following recipe: Let $n_{i,j} = y_{i,j} - \hat{y}_{i,j}$ be the data residual at transect i and survey j. We remove the mean of the residuals at each survey from all the residuals at that survey, so that $\Sigma_i n_{i,j} = 0$. The autocorrelation at lag k is then estimated by $c_k = I^{-1} J^{-1} \Sigma_j \Sigma_i n_{I+k,j} n_{i,j}$ scaled to have value 1 at zero lag. Figure 5 shows the autocorrelation data for Waihee, Maui, and the best-fit exponential decay for various lags. The expression for c_k shows that long lags are underweighted compared with short lags; and the fit was done using only data out to the first zero of the autocorrelation. Using the estimate for correlation distance, we build an estimated covariance matrix $\tilde{C}_{dd}$ that is block diagonal with one block for each survey. The time blocks are scaled by the *a priori* uncertainties for each survey, but each time block is proportional to $\exp(-|x_i - x_j|/L)$, where x_i and x_j are transect locations, and L is the correlation distance. We used this scheme because it does not require iteration and does not depend on the basis functions.

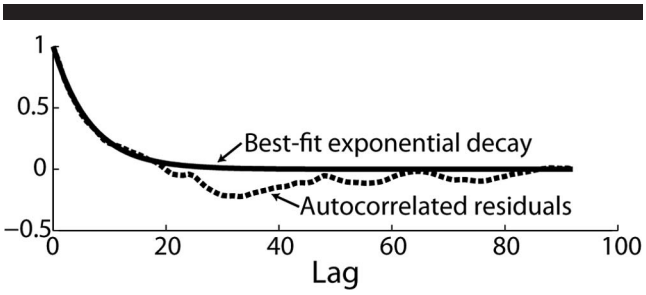

Figure 5. Autocorrelation of residuals (dashed line) and best-fit decaying exponential (solid line).

CALCULATIONS

Currently, two techniques are used in computing shoreline change rates along a beach—rates from single transects and rates from binned transects. Both techniques utilize shoreline data sets generated from topographic surveys (NOAA T-sheets and LIDAR surveys) and vertical aerial photographs.

Single-Transect Method

Most coastal scientists calculate shoreline change rates at each transect along a beach; for example, the Digital Shoreline Analysis System (DSAS) currently calculates shoreline change in this way (Thieler *et al.,* 2005). With the use of ArcGIS, DSAS creates transects that are separated by a distance defined by the user. At each transect, shoreline change rates from different statistical methods (*e.g.,* end-point rate, least squares regression) and their uncertainties are calculated. For each transect, three parameters are estimated from the data: slope, intercept, and variance; hence the total number of parameters is three times the number of transects. If the correlation distance of the alongshore noise exceeds the transect separation, single-transect overfits the data. Figure 6 shows the rates for various methods, including single transect.

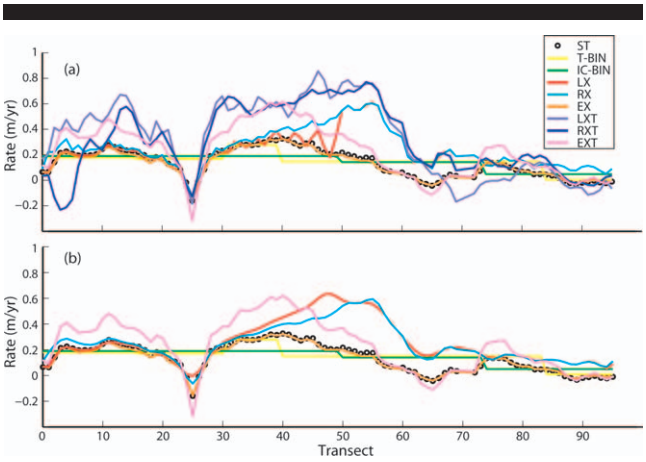

Figure 6. Change rates from various methods, Waihee, Maui. Positive change rates indicate erosion. In panel (a) the basis function sums for LX, LXT, RX, and RXT are truncated abruptly (boxcar window in the transform domain), while in panel (b) the best fit coefficients for those methods are modulated by half of a Hanning window. Rates shown for models with acceleration (EXT, LXT and RXT) are those at the time of the most recent survey, while rates for constant-rate models (EX, LX, RX, t-bin, IC-bin) are centered at $\bar{t}$. Rates found with the single-transect method (open circles) are significant for AICm and gMDL, but not for other ICs. The polynomial methods (LX, LXT, RX, RXT) struggle to fit the spike in local accretion near transect 25 because spikes and jumps in rate are only weakly identifiable with smooth basis functions. IC-bin favors a three-bin model under all ICs but gMDL. EXT gives the best fit to data (see Table 1) followed closely by EX, which agrees best with single-transect.

Binned Transects

Genz *et al.* (2007) introduced binning as a technique that reduces the effects of data scatter by identifying adjacent transects that have indistinguishable rates, and grouping them together into cells (bins) with a single rate. The resolution of the rate is not as high as with single transects, but the error in the change rate estimate is reduced within each bin compared with single-transect. Binning methods give rates that are discontinuous at the bin boundaries, but this is useful if a beach is to be divided into distinct management zones.

The forward model for beach erosion is

$$y_{ij} - \bar{y}_i = r(t_j - \bar{t}) + n_{ij}; \qquad x_i \in \text{bin}$$

in which n_{ij} is a noise sample at time t_j, and the rate r is the same for each transect in the bin. As noted previously, our baseline shift has made it unnecessary to include an intercept. The inversion for bin boundaries is by search, but for a given set of bin boundaries, we use GLS to find the rate within each bin. In Equation (5) the degrees of freedom count is $\nu = T_b(J - 1) - 1$, where T_b is the number of transects in the bin, and J is the number of surveys. We use formula (1a) to calculate an IC for the whole beach, setting N equal to the total number of data points in the data set, and K equal to the number of transects (one intercept per transect), plus twice the number of bins (a rate and variance for each bin) plus one for the correlation length parameter.

The binning technique of Genz *et al.* (2007) was based on the t-test for independence of adjacent change rates, and it required significant user input. In the following discussion, we give an improved binning technique that is less subjective and easier to use. Hereinafter, we refer to Genz's *et al.* (2007) technique as "t-binning," for the t-test, and the improved technique of this paper as 'IC-binning', for the IC score used to select the bins. Any IC can be used to select the bins.

t-Binning

t-Binning uses Student's t-test statistic (Kleinbaum *et al.*, 1998) to determine whether a binned group rate is statistically significant. With a bin-window length of four transects, the first four adjacent transects are grouped and a t-test is used to determine whether their change rates are significantly different. The window then shifts by one transect, and the transects in the new window are used to calculate a change rate and a t-test statistic. The window keeps shifting until it reaches the end of the beach. Then the window spacing is increased to six transects and the process is repeated. This is done until the window spacing equals half the number of transects. Once complete, transect groups with a significant t-test at each window size are displayed. Another t-test is computed on significant transect groups within each window size. Based on the t-test results, the user identifies clusters of transects within each window. The next step is to compare clusters of transects in one window size to overlapping clusters of other window sizes. This step also utilizes the t-test, and the user makes the final decision of how the transects are grouped together. t-Binning is not practical for long stretches of beach, as varying t-test results make it difficult to identify bins. For Waihee, Maui, the best-fit t-bin model had five bins, as shown in Figure 6, and $K = 104$.

IC-Binning

IC-binning uses scores from an IC to identify bins, and it is considerably simpler than t-binning. The one-bin model has one rate for the whole beach. For a beach with I transects, there are $I - 1$ possible two-bin models corresponding to the $I - 1$ choices for the bin boundary, and each two-bin model has a rate for each bin. Continuing in this way, one arrives at the I-bin model, in which each transect has its own bin; this model is the single-transect model, now seen to be a special case of binning. IC-binning systematically calculates an IC score for each bin configuration, then chooses the configuration with the lowest score. As models are examined in order of increasing number of bins, the IC score first declines then begins to rise as the number of bins begins to overfit the data. As soon as the IC score begins to rise, we stop examining models, and the model with the lowest score is chosen. IC-binning is computationally intensive because of the large number of candidate models, although the computation for each model is relatively rapid. For Waihee, Maui, AICc, AICm, AICu, and SIC favored a three-bin model (Figure 6) for which $K = 100$, while gMDL favored a four-bin model ($K = 103$) over a three-bin model. Only models up to four bins were examined because of long execution times. For beaches

this size or larger, IC-binning should be powered by a Gibbs sampler or a genetic algorithm.

Polynomial Methods

In all the methods of this paper, the shoreline change rate is represented as a linear sum of basis functions. In the binning method, each basis function is a unit "boxcar," equal to unity within its bin and vanishing elsewhere; in the single-transect method each boxcar narrows to a delta function. By contrast, in the polynomial methods the basis functions are linear sums of powers of x. In the LX method (see next section), the first basis function represents the average rate of the entire beach. In the single-transect method, the basis functions are as local as possible whereas in the LX and RX methods the basis functions are global.

Similar to binning, the polynomial technique reduces the number of parameters. However, unlike binning, the rates vary smoothly along the beach, eliminating the discontinuities that result from binning and single-transect. Rates produced by the polynomial technique can have lower error than rates produced by binning models, although this depends on the particular shoreline—a shore with sharp littoral cells might be best fit by a binning model. We refer to models in which rates are assumed time-independent as PX models (for polynomials in x) and to models in which rates can change with time as PXT models (for polynomials in x and t). LX and LXT mean that Legendre polynomials are used, while RX and RXT mean that trigonometric functions (sines and cosines) are used.

LX and LXT

Because LX is a special case of LXT, we present only the latter. For LXT, the parameter vector consists of coefficients of Legendre polynomials (Weisstein, 1999) sampled at the transect locations. The forward model for beach erosion contains an intercept sum, a rate sum, and an acceleration sum:

$$y_{ij} - \bar{y}_i = \sum_{k_0=0}^{K_0} \alpha_{k_0}^{(0)} P_{k_0}[z(x_i)] + (t_j - \bar{t}) \sum_{k_1=0}^{K_1} \alpha_{k_1}^{(1)} P_{k_1}[z(x_i)] + \frac{1}{2}(t_j - \bar{t})^2 \sum_{k_2=0}^{K_2} \alpha_{k_2}^{(2)} P_{k_2}[z(x_i)] + n_{ij},$$

where the $P_k(z)$ are Legendre polynomials on the interval $[-1, 1]$, and the n_{ij} are samples from the noise affecting data at time t_j. The function $z(x)$ maps the interval (x_1, x_1) to the interval $(-1, 1)$ by $z(x) = -1 + 2(x - x_1)/(x_1 - x_1)$.

The terms in the LX model have straightforward physical interpretations. In the rate sum, *i.e.,* the sum with factor $(t_j - \bar{t})$, the average rate of shoreline change is the coefficient of P_0 because $P_0(x) = 1$. Similarly, because $P_1(x) = x$, the coefficient of P_1 in the rate sum is simply related to the flux of the area from one end of the beach to the other. Roughly speaking, in the rate sum, transfers of amplitude from terms of even degree to terms of odd degree represent transfer of area in the alongshore direction. In the acceleration sum (the third sum in the model), the coefficient of P_0 gives the average acceleration of shoreline change.

The LX model is obtained as a special case of LXT by omitting the acceleration terms and (because of the preprocessing baseline shift) the intercept terms; *i.e.,* the first and third summations in the LXT formula are omitted. When these terms are present for LXT, we fix the index K_0 equal to K_2. If transects are not evenly spaced along a beach, the sampled Legendre polynomials are not necessarily orthogonal, and we orthonormalize them with the Gram-Schmidt procedure.

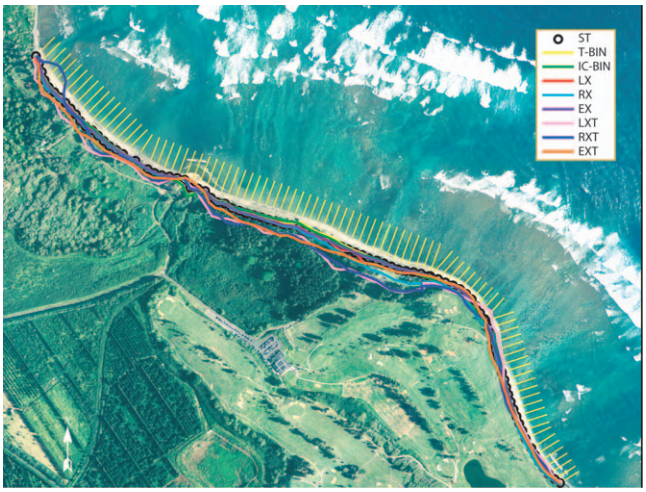

Figure 7. Fifty-year hazard prediction from various methods, Waihee, Maui.

A mathematically inclined reader may wonder why Legendre polynomials were used instead of the Chebyshev polynomials $T_n(x)$ because the latter more closely resemble sines and cosines. The answer is that Legendre polynomials are orthogonal in the usual inner product, whereas each Chebyshev polynomial has a dual function that is singular at $x = \pm 1$. Accordingly, a matrix whose columns consist of sampled Chebyshev polynomials may have a poorly behaved inverse. For a local analysis at one end of a beach, Laguerre polynomials can be used, and for a local analysis in the interior of a long beach, Hermite polynomials can be used. Coding details for LXT are given in the mathematical version of this paper available from the authors.

To calculate an IC score for the polynomial method, we use Equations (1) with $K = I + K_1 + 2K_2 + 2$. Because there are I transects, it follows that $0 \leq K_1, K_2 \leq I$, and the number of possible LXT models is I^2. If we had admitted basis function sums with missing lower-order terms, the number of possible models would have been 2^{2I}, and the IC scores of LXT might have been lowered further. For Waihee, Maui, the best-fit LX model had $K = 174$, and the best-fit LXT model had $K = 242$. Figure 7 shows the 50-year prediction of the shoreline, and Figure 8 shows the residuals for all methods.

RX and RXT

Legendre polynomials are a reasonable choice for most beaches because $P_0(x) = 1$ and $P_1(x) = x$. However, for island shorelines without endpoints, sines and cosines can give mod-

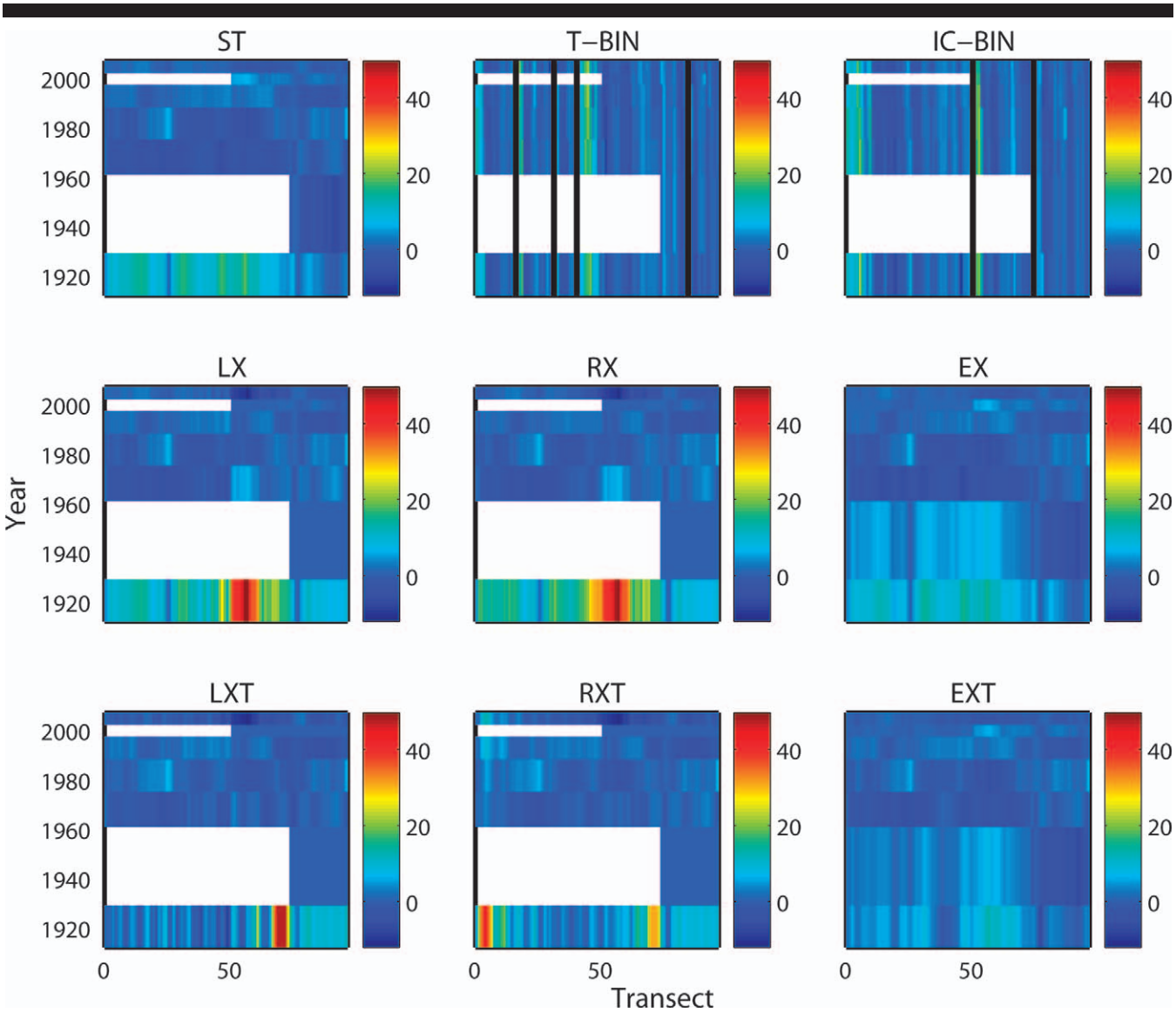

Figure 8. Residuals in meters for various methods, Waihee, Maui. Black lines on the residuals for t-bin and IC-bin indicate bin boundaries. White areas indicate missing data, except for EX and EXT, which interpolated missing data. Inclusion of acceleration improves the fit to the 1920 survey. EX and EXT give the best fits to data.

els with lower IC scores. To use trigonometric functions instead of Legendre polynomials, we make the substitutions

$$P_n[z(x)] \rightarrow \begin{cases} \cos\left[\dfrac{n+0}{2}\pi z(x)\right] & \text{if } n = 0, 2, 4, \ldots \\ \sin\left[\dfrac{n+1}{2}\pi z(x)\right] & \text{if } n = 1, 3, 5, \ldots \end{cases}$$

In other words, if n is even, $P_n(z(x))$ is replaced by $\cos[n\pi z(x)/2]$ and if n is odd, $P_n(z(x))$ is replaced by $\sin[(n+1)\pi z(x)/2]$. This correspondence was chosen so that the substituted trigonometric functions are roughly similar in shape to the Legendre polynomials they replace. As noted in the preceding section, we refer to the resulting method as RXT, and the version without acceleration as RX. For Waihee, Maui, the best-fit RX model had $K = 179$ and the best-fit RXT model had $K = 242$.

Eigenbeaches (EX and EXT)

In the previous section we expanded the rate and acceleration in preselected polynomials, but in this section we look for a different set of basis functions based on the beaches themselves. To make this notion more precise, suppose that the linearized beach physics is approximately described by $dy/dt = Ay + f$. Here, as usual, y is an $N \times 1$ column vector with component $y_i(t)$ representing the shoreline at alongshore

location x_i relative to the baseline $\bar{y}_i$. The beach operator or "beach process" is an $N \times N$ matrix, A, and the source term $f(t)$ is another $N \times 1$ column vector. In many physics problems, A has the decomposition $U\Lambda U^T$, where the columns of U are the eigenvectors of A, and the matrix Λ is diagonal. It follows that

$$y(t) = U\left[e^{\Lambda(t-t_1)}U^T y(t_1) + \int_{t_1}^{t} dt' e^{\Lambda(t-t')}U^T f(t')\right],$$

a result that is easy to verify by differentiation. We do not need to compute anything with this equation; it is given here only to show that the beach at any given time is a linear sum of the eigenvectors of A. The coefficients in the linear sum are the rows of the column vector $[\ldots]$ on the right-hand side of the expression for $y(t)$. To find the eigenvectors, we collect successive beaches as columns of a matrix, subtract the mean column from each column, and then find the principal components of this *beach matrix*. If our conception of the beach physics is even approximately correct, the principal components of the beach matrix will give a more parsimonious description of the change process than can be gotten with polynomials, so it makes sense to use them as basis functions. We refer to these principal components as eigenbeaches.

We use eigenbeaches for prediction in the same way that we used Legendre polynomials previously. The first eigenbeach takes the place of the function $P_0(z(x_i))$; the second eigenbeach takes the place of the function $P_1(z(x_i))$, and so forth up to $P_p(z(x_i))$. In the polynomial method, the summation upper limits K_0, K_1, K_2 cannot exceed the number of transects I, but in the eigenbeaches method, these upper limits cannot exceed the number of surveys J. As for most data sets $J \ll I$, there is a strong suggestion that fewer terms are needed in each sum than in the polynomial method, and this turns out to be the case. Eigenbeaches with acceleration is referred to as EXT, and without acceleration, as EX.

In the EX/T method, because the basis functions come from the data, it is inappropriate to use the same data to compute the RSS term in an IC. We therefore divide the original data set into two data sets. We use the first data set to generate basis functions, then use those basis functions to model the second data set. We repeat this procedure, interchanging the data sets. For our final model, we use all the data at once to generate the basis functions, but the allowed number of rate basis functions, K_1, is related to the number of rate basis functions from the two earlier runs by choosing the integer nearest to $2^{-1/2}(K_1^{(1)} + K_1^{(2)})$ with a similar formula for acceleration basis functions. The logic behind this procedure is that signal-to-noise ratio increases roughly as the square root of the number of data. Figure 9a shows the eigenbeaches for Waihee Maui, and Figure 9b shows the time variation in the coefficients. For Waihee, the best-fit EX model had $K = 97$, and the best-fit EXT model had $K = 100$.

While this paper was in review, two papers appeared that use methods related to our EX method: Miller and Dean (2007a, 2007b) used empirical orthogonal functions (EOF) to characterize modes of shoreline variability. Miller and Dean (2007a) used the characteristic shape of their spatial EOFs to develop parametric and nonparametric basis functions tailored to particular shorelines, and Miller and Dean (2007b) related modes of shoreline variability to physical parameters in the nearshore environment. Because their interest was in variability rather than long-term prediction, they used time EOFs as well as space EOFs. Because our interest is long-term prediction, our time functions are simple rate and acceleration terms. To understand how the two approaches can be combined, imagine a data set with dense time samples in a 5–10 year interval, and sparse time samples in a 50-year interval. With a reasonable stationarity assumption, the samples in the 5–10 year interval can tell us about the noise in the 50-year interval. If we orthogonalize the 5–10 year spatial EOFs with respect to the 50-year EOFs, the 5–10 year EOFs can be used to estimate the noise covariance for the long term process, leading to more accurate long term predictions and confidence intervals. Thus, denser, modern data multiply the value of older sparser data.

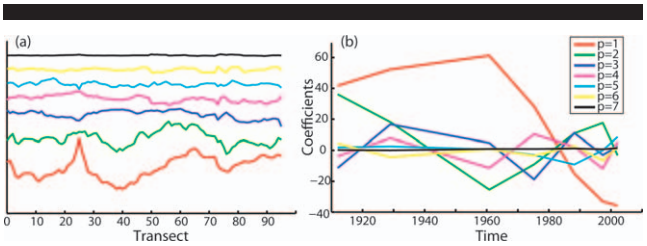

Figure 9. (a) Eigenbeaches scaled by eigenvalue for Waihee, Maui. (b) Eigenvalue coefficients *vs.* time. The coefficient of the eigenbeach with largest eigenvalue shows the clearest trend with definite acceleration.

Robustness

As promised in discussion at the end of the Introduction, we outline how to make the methods of this paper robust to violations of the additive Gaussian noise assumption. Dramatic examples of the need for robustness are shown in figures 1–3 of Douglas and Crowell (2000), and in figure 4 of Genz *et al.* (2007). Of course, unusual data points associated with large storms can be removed from data sets, but the effects of storms and the duration of those effects are not always known; moreover, the editing of data introduces an undesirable subjectivity into the analysis. Robust methods include least absolute differences (LAD), and least median of squares (LMS). In LAD processing—which has a very long history, beginning with Laplace, and is sometimes referred to as L_1 processing, as opposed to L_2 (least squares) processing—the generalized least squares likelihood function is replaced by

$$2^{-N/2}|C_{dd}|^{-1/2}\exp[-\sqrt{2}\|C_{dd}^{-1/2}(d - Gm)\|_1],$$

in which $\|\ldots\|_1$ is an instruction to sum of the absolute values of components. Because the covariance matrix is symmetric and positive definite, its positive square root is easily computed. This likelihood function does not have continuous derivatives, so one must find the optimal model by searching, using either enumeration (which is best up to 5–6 parameters), or by Markov chain Monte Carlo methods (MCMC) such as Gibbs sampling (*e.g.,* Basu and Frazer, 1990; Brooks and

Frazer, 2005; Dosso, 2002; Sambridge and Mosegard, 2002). Models with 10–20 parameters are not onerous for today's desktop computers, and MCMC methods have advantages not available to least squares. For example, they allow easy incorporation of prior information, even if the prior is not conjugate, and they can be used with other information criteria, such as the deviance information criterion (Spiegelhalter *et al.,* 2002).

With regard to use of an IC in LAD processing, Hurvich and Tsai (1990) derived an IC under the assumption of Laplace (double exponential) noise; in tests on synthetic data with Laplace noise, they found that their L_1 version of AICm did not perform better than the usual L_2 version of AICm, and it took significantly longer to compute. Accordingly, the use of L_2 ICs on data with L_1 noise does not appear to be an issue of concern.

In least median of squares (LMS), one looks for the parameter vector that minimizes the median of the squared residuals instead of the average of the squared residuals (Rousseuw, 1984; Rousseuw and Leroy, 1987). Data points identified as outliers by the LMS processor are then discarded, and the remaining data are analyzed with (in our case) generalized least squares. Rousseuw and Leroy (1987) give a number of examples in which LAD breaks down and LMS does not. If a storm is known *a priori* to be fairly uniform in intensity along a shoreline, it may be appropriate to omit an entire survey rather than simply omitting points that do not agree with the trend. For the methods of this paper, the LMS processor can use the same search techniques as LAD. Many Hawaii beaches have been processed with LMS-RLS with good results (Fletcher *et al.,* 2003; Rooney, 2002; Rooney and Fletcher, 2005). In their comparison of LAD and LMS methods with other methods for single transect processing, Genz *et al.* (2007) found that LAD and LMS were both superior but that neither was notably superior to the other if there was only one storm point present in the data. Figure 4 of Genz *et al.* (2007), has several storm points, and it would not have been processed correctly by LAD. In the context of another problem, Antille and El May (1992) give an example in which LMS breaks down but LAD does not. Birkes and Dodge (1993) give a readable introduction to both LAD and LMS.

DISCUSSION

Moving Windows

A reviewer of this paper asked: Why not analyze the data in a moving window? To address this question, note that a moving window procedure resembles our IC-binning method, but with a sliding bin; the rate output at each x would be calculated using data in a window centered at that x. More sophisticated moving-averaging schemes address this difficulty by allowing the rate to vary smoothly within the moving window. In *local polynomial regression,* for example, the data in the window are fit to a polynomial, and an IC is used to determine the order of the polynomial, just as we do here for the whole data set. The parsimony argument against all such methods is that analyzing a shoreline with 100 transects in this way still delivers 100 change rates, just as in a single-transect analysis. By comparison, our LX and EX procedures parse shoreline change into a small number of physically meaningful components. In the language of physics, they reduce the dimension of the configuration space. Reduction of dimension nearly always leads to increased understanding and better predictions, especially in nonlinear systems.

Windowing in the Transform Domain

It is often useful to regard basis function methods as a kind of filtering. For example, if the sampled Legendre basis functions are orthogonalized (we do this with the Gram-Schmidt procedure) and if noise is uncorrelated, then LXT consists of a Legendre transform of each survey, followed by regression of the coefficients of like degree to the time functions $\{1, (t - \bar{t}), \frac{1}{2}(t - \bar{t})^2\}$, then truncation in degree, then inverse transformation. From this point of view, the function of an IC is to tell us how long to make the boxcar window applied to the data in the transform domain. However, to avoid ringing over alongshore discontinuities in rate, one can replace the boxcar with, say, half of a Hanning window (Bracewell, 2000), as in Figure 6(b). EX, it should be noted, does not ring across alongshore discontinuities in rate because such discontinuities will be represented in the lowest order eigenbeach; however, truncation still leads to artifacts. In using a boxcar window, we implicitly assume that use of an IC keeps truncation artifacts below the level of noise, which might be the case for strongly identifiable models but is unlikely to be the case for weakly identifiable models. Notably, the use of an IC provides a natural alternative to the tapering of coefficients, which is to weight the predictions of different order models by their relative IC-probabilities $p_i \propto \exp(-\mathrm{IC}_i/2)$. Such extensions to our procedures raise questions that are beyond the scope of this paper but will be explored elsewhere.

Noise

Our treatment of noise is as simple as possible to compare different methods in an uncomplicated way. We assume that noise is additive and Gaussian, that the scale length of the noise process is the same at each transect location, and that the noise in any one survey is uncorrelated with the noise in other surveys. We also assume that noise power is independent of transect, and that the relative levels of noise in different surveys are known. Except for the additive noise, all of those assumptions can be relaxed at the cost of greater computation time. We estimated alongshore correlation length from the residuals of the single-transect method, then used that correlation length estimate with the other methods. In theory, results can be improved by estimating noise iteratively from the residuals for each method—ideally, the best model is the one that causes the data residuals to be uncorrelated in space and time—but our experiments with iteratively determining noise were inconclusive, possibly because of confounding by weakly identifiable models. Our handling of noise in the EXT method also raises questions because noise contaminates the basis functions themselves, an effect that we attempted to minimize by using cross-validation to compute the residuals in the IC formulas.

As noted in the preceding discussion, a GLS problem is transformed to an OLS problem by multiplying the data vec-

tor by the inverse square root of the noise covariance matrix. We elected to work with untransformed data partly because of questions such as: In the binning method, should bins be selected with transformed data or untransformed data? Multiplication by the inverse covariance matrix amplifies small spatial wavelengths and might result in a best-fit model with more bins than would be found with untransformed data. In retrospect, such questions now seem to us to be addressable independent of the transform from GLS to OLS, and it is mainly a matter of correct coding.

CONCLUSIONS

In any survey, it is always desirable to have as many transects as possible to reduce the effects of noise. However, transects that are close together are not independent, and so the traditional single-transect model leads to overparameterization of the data. Here we have introduced three new classes of methods for predicting beach erosion: IC-binning, polynomial methods, and eigenbeaches. These methods all utilize basis functions, and the traditional single-transect method can be regarded as a limiting member in which each basis function is a delta function. The use of more general basis functions leads to parsimony via a reduction in the number of model parameters compared with the traditional single-transect method. The reduced number of parameters is a matter of practical importance because it occasionally allows one to detect acceleration or to make a statistically defensible prediction with a data set that seems hopelessly noisy when transects are independently analyzed. IC-binning, which includes the single-transect method as a special case, forces rate to be piecewise constant along the beach, which may be useful for shoreline managers. The polynomial methods parse patterns of change into components that are easily interpreted, and they force rate to vary smoothly along the beach, which may be appropriate when the rate is to be used for long-term prediction. The eigenbeaches method avoids assumptions about rate smoothness by generating its basis functions from the data themselves. These methods are not exhaustive—imagine piecewise LX—but they illustrate the possibilities. Our tests of various ICs support the much more extensive work of McQuarrie and Tsai (1998) in finding that AICu is a dependably good IC, especially when sample size is less than 100. For sample sizes of several hundred or more, we found that SIC also did well. For a given shoreline, the difference between good ICs like AICu and SIC is probably much less important than the difference between single-transect, binning, PXT, and EXT.

In summary, we favor LXT for short, smooth shorelines without hardening, and EXT for long shorelines and shorelines with hardened sections. EXT seems to us to be the method most worthy of further development. The single-transect method should always be run first and compared with other methods, even when the rates it gives are not significant. If different methods agree, a prediction is more likely to be meaningful, and if they do not agree, it is prudent to try to understand why.

ACKNOWLEDGMENTS

Funding for this study was provided by the Coastal Marine Geology Program, National Shoreline Assessment Project of the U.S. Geological Survey, the University of Hawaii Sea Grant College of the National Oceanographic and Atmospheric Administration, the Hawaii Department of Land and Natural Resources, Maui County Planning Department, Kauai County Planning Department, U.S. Army Corps of Engineers, Hawaii Coastal Zone Management Program, and the City and County of Honolulu. We thank the University of Hawaii Coastal Geology Group for their assistance.

APPENDIX A. AICc WITH CORRELATED NOISE

Here we briefly justify the IC formulas used in the body of the paper. A reader who has never encountered ICs before might first want to read the derivation in Burnham and Anderson (2002, p. 362) or Linhart and Zucchini (1986, p. 243). All IC formulas are based on likelihood and penalty, and we extend them to correlated data by using the likelihood function for correlated data. Practically speaking, the net effect of the extension is to change the definition of RSS that appears in the IC formulas. The new definition of RSS reduces to the old one when the data are uncorrelated. The extension is given here for AICc, and the extensions of the other ICs are obvious.

The likelihood of data parameter vector m, given data d, is defined to be the probability density function of d given parameter vector m. Our model is $d = Gm + n$, where n is a Gaussian noise process with covariance matrix C_{dd}, so the likelihood function is

$$L(m\,|\,d) = (2\pi)^{-N/2}|C_{dd}|^{-1/2}\exp\left[-\frac{1}{2}(d - Gm)^T C_{dd}^{-1}(d - Gm)\right].$$

We also need the quantity LL, defined to be -2 times the logarithm of the likelihood,

$$LL = N\log(2\pi) + \log|C_{dd}| + (d - Gm)^T C_{dd}^{-1}(d - Gm).$$

Although $\hat{C}_{dd}$ is a matrix, it is customary to regard it as a single parameter, and this is done by assuming that $\tilde{C}_{dd}$ is proportional to some *a priori* estimate $\tilde{C}_{dd}$ with constant of proportionality α. Thus, α is the variance parameter to be estimated.

Intuitively it is obvious that models with low LL (high likelihood) fit the data better than models with high LL (low likelihood). A first version of the AICc formula is obtained by adding a penalty term to LL evaluated at maximum likelihood values,

$$\text{AICc} \approx \text{LL}(\hat{m}, \hat{\alpha}) + 2KN/(N - K - 1),$$

in which $\hat{m}$ and $\hat{\alpha}$ are the values that minimize LL, and the penalty term reduces the bias caused by having more parameters in the model (Hurvich and Tsai, 1989; Sugiura 1978).

To find $\hat{m}$ and $\hat{\alpha}$, we differentiate LL with respect to m and α, then set the derivatives to zero, obtaining $\hat{m} = (G^T\tilde{C}_{dd}^{-1}G)^{-1}G^T\tilde{C}_{dd}^{-1}d$, and $\hat{\alpha} = N^{-1}(d - G\hat{m})^T\hat{C}_{dd}^{-1}(d - G\hat{m})$, the maximum likelihood estimators (MLE) of m and α, respectively. Putting the MLEs into the AICc formula gives

$$\mathrm{AICc} \approx N + N\log(2\pi) + N\log[N^{-1}|\tilde{C}_{dd}|^{1/N}(d - G\hat{m})^T \times \tilde{C}_{dd}^{-1}(d - G\hat{m})] + 2KN/(N - K - 1).$$

The determinant property $|aA| = a^N|A|$ shows that this result does not change if $\tilde{C}_{dd}$ is multiplied by any positive number. Because the number of data points N is common to all terms in the formula, and because N does not depend on method, we divide by N. Finally, because the first two terms in AICc are the same for every model, we omit them from the formula, obtaining Equation (1a).

APPENDIX B. gMDL WITH CORRELATED NOISE

In the nomenclature of Hansen and Yu (2001, 2003), the version of MDL used by Fenster, Dolan, and Elder (1993) is a modified SIC criterion (Hansen and Yu refer to it as BIC), and the SIC is a two-stage MDL. From Hansen and Yu (2001, 2003) we adapt the mixture MDL they refer to as gMDL. Our gMDL statistic is

$$\mathrm{gMDL} = \begin{cases} \log S + \dfrac{K}{N}\log F + \dfrac{2}{N}\log N, & \text{if } R^2 \geq \dfrac{K}{N} \\ \log\dfrac{\mathrm{DSS}}{N} + \dfrac{1}{N}\log N, & \text{otherwise.} \end{cases}$$

To compute the gMDL formula, it is helpful to code the following quantities:

$\mathrm{DSS} = |\tilde{C}_{dd}|^{1/N} d^T \tilde{C}_{dd}^{-1} d$, the sum of squared data,

$\hat{d} = G(G^T\tilde{C}_{dd}^{-1}G)^{-1}G^T\tilde{C}_{dd}^{-1}d$, the predicted data

$\mathrm{RSS} = |\tilde{C}_{dd}|^{1/N}(d - \hat{d})^T\tilde{C}_{dd}^{-1}(d - \hat{d})$, the residual sum of squares

$\mathrm{FSS} = \mathrm{DSS} - \mathrm{RSS}$, the fitted sum of squares

$S = \mathrm{RSS}/(N - K)$, the S-statistic

$F = \mathrm{FSS}/(KS)$, the F-statistic

$\bar{d} = N^{-1}\sum_{i=1}^{N} d_i$, the mean of the data vector,

$\mathrm{MSS} = |\tilde{C}_{dd}|^{1/N}(d - \bar{d})^T\tilde{C}_{dd}^{-1}(d - \bar{d})$, the sum of squared deviations from the mean,

$R^2 = 1 - \mathrm{RSS}/\mathrm{MSS}$, the R-squared statistic,

$\tilde{C}_{dd} = \mathrm{diag}(\sigma_1^2 C_L, \sigma_2^2 C_L, \ldots, \sigma_J^2 C_L)$, the block-diagonal data covariance matrix,

σ_j^2: $1 \leq j \leq J$, the *a priori,* data-uncertainty scaling factors, one for each survey,

$(C_L)_{ij} = \exp(-|x_i - x_j|/L)$, the submatrix for each time block,

L, the correlation length of the spatial noise process.

In mixture MDL a posterior probability density function (pdf) of m is analyzed instead of a likelihood function and the prior is a conjugate prior whose parameters are referred to as hyperparameters. In the MDL formula just given, the hyperparameters are hidden because their values have been selected using maximum likelihood (Hansen and Yu, 2001, 2003). Adjustment of a prior pdf using data violates the Bayesian spirit of MDL, so mixture MDL is likely to undergo future theoretical refinement. In the name gMDL, the "g" comes from Zellner (1986) who noted that mixture MDL was simplified computationally by use of a prior for m that has a covariance matrix proportional to $(G^TG)^{-1}$; Zellner (1986) called this a g-prior. Our adaptation of gMDL to correlated noise assumes a prior for m with covariance matrix proportional to $(G^T\tilde{C}_{dd}^{-1}G)^{-1}$.

LITERATURE CITED

Antille, G. and El May, H., 1992. The use of slices in the LMS and the method of density slices: foundation and comparison. *In*: Dodge, Y. (ed.), *Computational Statistics.* Heidelberg, Germany: Physica-Verlag.

Basu, A. and Frazer, L.N. 1990. Rapid determination of critical temperature in simulated annealing inversion. *Science,* 249, 1409–1412.

Birkes, D. and Dodge, Y. *Alternative Methods of Regression.* New York: John Wiley.

Bracewell, R.N., 2000. *The Fourier Transform and Its Applications,* 3rd edition. Boston: McGraw-Hill.

Brooks, B.A. and Frazer, L.N., 2005. Importance reweighting reduces dependence on temperature in Gibbs samplers: an application to the coseismic geodetic problem. *Geophysical Journal International,* 161(1), 12–20.

Burnham, K.P. and Anderson, D.R., 2002. *Model Selection and Multimodel Inference,* 2nd edition, New York: Springer-Verlag.

Crowell, M.; Douglas, B.C., and Leatherman, S.P., 1997. On forecasting future U.S. shoreline positions: a test of algorithms. *Journal of Coastal Research,* 13(4), 1245–1255.

Dosso, S., 2002. Quantifying uncertainties in geophysical inversion I: a fast Gibbs sampler approach. *Journal of the Acoustical Society of America,* 111, 129–142.

Douglas, B.C. and Crowell, M., 2000. Long-term shoreline position prediction and error propagation. *Journal of Coastal Research,* 16(1), 145–152.

Draper, N.R. and Smith, H., 1998. *Applied Regression Analysis,* 3rd edition. New York: John Wiley.

Fenster, M.S.; Dolan, R., and Elder, J.F., 1993. A new method of predicting shoreline positions from historical data. *Journal of Coastal Research,* 9(1), 147–171.

Fletcher, C.H.; Rooney, J.J.B.; Barbee, M.; Lim, S.-C., and Richmond, B.M., 2003. Mapping shoreline change using digital orthophotogrammetry on Maui, Hawaii. *Journal of Coastal Research,* Special Issue No. 38, 106–124.

Galgano, F.A. and Douglas, B.C., 2000. Shoreline position prediction: methods and errors. *Environmental Geosciences,* 7(1), 23–31.

Genz, A.S.; Fletcher, C.H.; Dunn, R.A.; Frazer, L.N., and Rooney, J.J., 2007. The predictive accuracy of shoreline change rate methods and alongshore beach variation on Maui, Hawaii. *Journal of Coastal Research,* 23(1), 87–105.

Hansen, M.H. and Yu, B., 2001. Model selection and the minimum description length principle. *Journal of the American Statistical Association,* 96(454), 746–774.

Hansen, M.H. and Yu, B., 2003. Minimum description length model selection criteria for generalized linear models. *Institute for Mathematical Statistics Lecture Notes,* Monograph Series, 40, 145–164.

Honeycutt, M.G.; Crowell, M., and Douglas, B.C., 2001. Shoreline-position forecasting: impact of storms, rate-calculation methodologies, and temporal scales. *Journal of Coastal Research,* 17(3), 721–730.

Hurvich, C.M. and Tsai, C.-L., 1989. Regression and time series model selection in small samples. *Biometrika,* 76, 297–307.

Kleinbaum, D.G.; Kupper, L.L.; Muller, K.E., and Nizam, A., 1998.

Applied Regression Analysis and Other Multivariable Methods, 3rd edition. Pacific Grove, CA: Duxbury Press.
Linhart, H. and Zucchini, W., 1986. *Model Selection.* New York: John Wiley.
McQuarrie, A.D.R., and Tsai, C.-L., 1998. *Regression and Time Series Model Selection.* Singapore, World Scientific.
Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* Revised edition. San Diego, CA: Academic Press.
Miller, J.K. and Dean, R.G., 2007a. Shoreline variability via empirical orthogonal functions analysis: Part I temporal and spatial characteristics. *Coastal Engineering,* (54), 111–131.
Miller, J.K. and Dean, R.G., 2007b. Shoreline variability via empirical orthogonal functions analysis: Part II relationship to nearshore conditions. *Coastal Engineering,* (54), 133–150.
National Academy of Sciences, 1990. *Managing Coastal Erosion.* Washington DC: National Research Council, Committee on Coastal Erosion Zone Management, National Academy Press.
Rissanen, J., 1978. Modeling by shortest data description. *Automatica,* 14, 465–471.
Rooney, J.J.B., 2002. *A Century of Shoreline Change along the Kihei Coast of Maui, Hawaii.* Manoa, Hawaii: University of Hawaii, Ph.D. dissertation, 174p.
Rooney, J.J.B., and Fletcher, C.H., 2005. Shoreline change and Pacific climate oscillations in Kihei, Maui, Hawaii. *Journal of Coastal Research,* 21(3), 535–547.
Rousseuw, P.J., 1984. Least median of squares regression. *Journal of the American Statistical Association,* 79(388), 871–880.
Rousseuw, P.J. and Leroy, A.M., 1987. *Robust Regression and Outlier Detection.* New York: John Wiley.
Sambridge, M. and Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics,* 40(3), 1009.
Schwarz, G., 1978. Estimating the dimension of a model. *Annals of Statistics,* 6, 461–464.
Shibata, R., 1980. Asymptotic efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics,* 8, 147–164.
Sugiura, N., 1978. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics, Theory and Methods,* A7, 13–26.
Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P., and Van der Linde, A., 2002. Bayesian measures of complexity and fit, *Journal of the Royal Statistical Society: Series B,* 64(4), 583–639.
Thieler, E.R.; Himmelstoss, E.A.; Zichichi, J.L., and Miller, T.L., 2005. Digital shoreline analysis system. http://woodshole.ber.usgs.gov/project-pages/dsas/ (accessed June 2006).
Weisstein, E.W., 1999. Legendre Polynomial http://mathworld.wolfram.com/LegendrePolynomial.html (March 2006).
Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *In:* Goel, P.K., and Zellner, A. (eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti.*, Amsterdam: North-Holland, pp. 233–243.
Zhang, K.; Douglas, B., and Leatherman, S., 2002. Do storms cause long-term beach erosion along the U.S. east barrier coast? *Journal of Geology,* 110(4), 493–502.