Long-term Shoreline Position Prediction and Error Propagation

Bruce C. Douglas and Mark Crowell

Department of Geography
University of Maryland
College Park, MD 20742, USA
bdouglas@bss2.umd.edu

Federal Emergency Management Agency
Washington, DC 20472, USA
mark.crowell@fema.gov

ABSTRACT


Beach erosion necessitates forecasting future shoreline positions for effective coastal management. Simple forecast methods, such as end-point rate and linear regression have been proposed in the coastal literature and are widely used. However, the matter of the error of forecasts has largely been neglected. If measurement errors and a linear trend of erosion were the only factors determining shoreline position, making predictions of future shoreline positions and their associated confidence intervals would be easy using linear regression. Unfortunately, real and sometimes enduring fluctuations of beach width occur that are much larger than the measurement uncertainty. Wintertime fluctuations of up to several 10's of meters are well-known; most investigators for this reason do not use winter shoreline positions to study long-term shoreline behavior. An individual great storm can cause beach erosion amounting to scores of meters requiring a decade or more for recovery. Using shoreline position data in linear regressions without considering storm-caused erosion and subsequent beach recovery may yield inaccurate predictions of future position resulting from the underlying erosion, and greatly inflated estimates of uncertainty (e.g., 95% confidence intervals). A case study of shoreline position change in Delaware is presented to show how consideration of knowledge other than shoreline positions alone can lead to useful results for shoreline position forecast errors. It is also demonstrated that modern, more accurate survey measurement techniques can be helpful in improving the quality of forecasts even if the inherent variability of shoreline position indicators remains at the level of many meters.

ADDITIONAL INDEX WORDS: Shoreline erosion rates, beach width fluctuation, forecast methods.

INTRODUCTION

GALGANO (1998) has shown that the vast majority (85–90%) of the US Atlantic coast is eroding. Long-term erosion rates vary from a small fraction of a meter per year, to erosion “hot spots” with rates of many meters per year. This variation of erosion rates has a number of causes, including geomorphic features such as inlets, wave energy, engineering changes, and others. Successful coastal management requires that long-term shoreline erosion rates be determined, and forecasts made of future shoreline positions along with estimates of their uncertainty.

It is common practice to base coastal construction setbacks on the erosion anticipated to occur over some specified time in the future, e.g., 30 or 60 years. Estimates of erosion rates can be made in a variety of ways (e.g., end point rate, linear regression), but are always subject to uncertainty because of measurement errors, and deficiencies in the model used to analyze historical shoreline position data. This uncertainty means that, to the setback predicted by the erosion rate, there must be added an additional quantity to ensure (to some level of confidence) that the property will be protected against erosion. However, calculated confidence intervals will depend on the assumed error of the measurements and the accuracy of the shoreline change model.

MORTON (1991) in a comprehensive review paper has discussed the complexity of shoreline position variation and the problem of interpreting the behavior due to the limited data sets available. FENSTER et al. (1993) proposed a methodology for analysis of shoreline position data and prediction of future shoreline position. The latter's application of the Minimum Description Length (MDL) algorithm was designed to provide a means for identifying erosion trend reversals, and provide a set of observation weights for making predictions using linear regression starting from the epoch of a trend reversal. However, they did not discuss for what period in the future their method would yield useful predictions, nor did they consider the matter of the errors of predicted positions.

Rapid recovery of beach width after ordinary winter storms is well known. After great storms, recovery can continue for as long as a decade (MORTON et al., 1994; GALGANO et al., 1998; DOUGLAS et al., 1998), so that any record of shoreline positions can have a highly irregular temporal pattern. CROWELL et al., (1997) noted that time series of sea level recorded by tide gauges display a trend plus quasiperiodic variability, and thereby resemble some series of shoreline positions. They tested the MDL and linear regression forecast methods using depleted samples of sea level data, and found that the data weighting scheme for the MDL algorithm proposed by FENSTER et al. (1993) did not yield consistently accurate results compared to a linear regression on the entire
data set. DOUGLASS et al. (1998) questioned the reliability of the linear model for predicting future shoreline locations, and suggested that confidence intervals be used to evaluate the quality of calculated erosion rates. They proposed that in mapping erosion hazard areas, sites might be considered non-erosional if the linear regression trends include zero recession at a selected confidence interval (the authors suggest an 80% confidence interval for their study area). However, DOUGLASS et al. (1998) confined their analysis to the issue of whether or not erosion was taking place at some level of confidence, and did not consider the issue of confidence intervals for forecast positions.

**METHODOLOGY**

Analysis of data begins with an attempt to find associations between variables, and regression analysis is the basic tool for discovering such associations. In the case of shoreline positions, a relation is sought between position and time. Linear regression can reveal if an association exists, and in particular (via the $r^2$ value) what fraction of the variance of the dependent variable (shoreline position) is attributable to the independent variable (time). What such a regression analysis cannot tell by itself is whether, for example, the linear regression model, consisting of a linear trend with added Gaussian random noise, is the appropriate one for the problem. However, new results (GALGANO et al., 1998; GALGANO, 1998; ZHANG, 1998) for shoreline position change since the middle of the 19th century are available for the US East Coast that indicate that the linear regression model is appropriate in certain cases, viz., shorelines unaffected by inlets or engineering changes. Such shorelines on US East Coast barrier islands appear to migrate landward with an underlying trend related to sea level rise (ZHANG, 1998), with additional superimposed punctuated episodes of erosion caused by great storms, followed by incremental multi-year accretion occurring during periods of storm quiescence.

We shall treat here the general case of fitting a polynomial of arbitrary degree in time, $t$, to both equally and unequally weighted shoreline position data sets. It is useful to consider the general case, because polynomial fits to shoreline data can sometimes be used (FENSTER et al., 1993) to identify reversals of shoreline trend. In addition, shoreline position measurements are becoming more precise because of the introduction of Global Positioning System (GPS) survey technology (BYRNES, 1993; LEATHERMAN, 1994), so that the case of unequal measurement weights must also be considered.

By hypothesis, the regression model is based on the assumption of a sequence of $k$ shoreline positions, $y_i$, linearly related to powers of time, $t$, by the relation

$$y_i = a + bt_i + ct_i^2 + \ldots + \text{noise}_i, \quad i = 1, \ldots, k \quad (1)$$

In matrix notation, this is compactly expressed as

$$Y = AX + N \quad (2)$$

where $Y$ is a $k \times 1$ column matrix, $A$ is a $k \times j$ matrix whose rows consist of $(1, t_i, t_i^2, \ldots)$, $X$ is a $j \times 1$ ($j =$ number of unknowns $a, b, c, \ldots$) column matrix, and $N$ is a $(k \times 1)$ column matrix of the noise. Assuming that the noise has zero mean, the weighted least squares solution to this problem in matrix notation is (GELB, 1974)

$$X = (A^TWA)^{-1}A^TY \quad (3)$$

where the superscript $T$ denotes transpose, and $W$ is the inverse of the covariance of the noise having dimension $(k \times k)$. For the usual assumption of statistically independent and normally distributed measurement errors of equal variance, we have

$$W = (1/(\sigma^2))I, \quad (4)$$

where $I$ is the identity matrix and $\sigma^2$ is the variance of the measurement noise. In this case, the $W$ matrices in equation (3) cancel, and the solution becomes

$$X = (A^TA)^{-1}A^TY \quad (5)$$

Carrying out the algebra indicated in Equation (5) for the case of two unknowns ($a$ and $b$ taken as intercept and slope) gives the familiar linear regression formulas.

The variance/covariance matrix of the estimated parameters is given by

$$\text{cov}(X) = (A^TWA)^{-1} \quad (6)$$

In the special case of equally weighted data as in equation (4)

$$\text{cov}(X) = \sigma^2(A^T)^{-1} \quad (7)$$

The diagonal terms in this matrix $\text{cov}(X)$ are the variances of the unknowns $a, b, c, \ldots$, and the off-diagonal terms are their covariances. Usually the latter are significant, and are needed to propagate the uncertainty of predicted values of the dependent variable (shoreline position in our case). The variance of estimated values ($Y$) of the dependent variable $y$ can be obtained from

$$\text{cov}(Y) = A \text{cov}(X)A^T \quad (8)$$

For the special case of linear regression with one independent and one dependent variable, these matrix equations collapse to simple forms, especially if the time epoch ($i.e.,$ origin) is taken to be the average of the observation times. For this special case, the matrix $A^TA$ has the following form for $n$ observations at times $t$, referred to the average time:

---

Table 1. Shoreline position data for Cotton Patch Hill, DE. Residuals are from a linear regression.

<table>
<thead>
<tr>
<th>Date</th>
<th>Position (m)</th>
<th>Type</th>
<th>Sigma* (m)</th>
<th>Residual** (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1845</td>
<td>0.0</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>18.2</td>
</tr>
<tr>
<td>1929</td>
<td>-84.8</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>-16.3</td>
</tr>
<tr>
<td>1944</td>
<td>-67.8</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>9.5</td>
</tr>
<tr>
<td>1954</td>
<td>-63.7</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>19.6</td>
</tr>
<tr>
<td>1962</td>
<td>-168.8</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>-80.8</td>
</tr>
<tr>
<td>1970</td>
<td>-116.1</td>
<td>NOS T SHEET</td>
<td>11.3</td>
<td>-23.3</td>
</tr>
<tr>
<td>1977</td>
<td>-86.9</td>
<td>AIR PHOTO</td>
<td>11.3</td>
<td>10.0</td>
</tr>
<tr>
<td>1990</td>
<td>-69.6</td>
<td>ORTHOPHOTO</td>
<td>7.0</td>
<td>35.1</td>
</tr>
<tr>
<td>1993</td>
<td>-101.2</td>
<td>GPS</td>
<td>7.0</td>
<td>5.3</td>
</tr>
<tr>
<td>1997</td>
<td>-85.8</td>
<td>GPS</td>
<td>7.0</td>
<td>23.9</td>
</tr>
</tbody>
</table>

SD of residuals = 35.4 M

* Includes 6.5 m assumed instability of the shoreline position indicator
** From a linear regression with all data equally weighted
Then $(A^t A)^{-1}$ follows by inspection, and a very simple formula for $\text{Cov}(\hat{Y})$ (i.e., the variance of an estimate of the independent variable, $y$) is obtainable. It is simply

$$\text{var}(\hat{Y}) = \sigma^2 \left[ \frac{1}{n} + \frac{1}{\sum (t_i)^2} \right].$$  \hspace{1cm} (9)

For the general case of unequally weighted data, it is necessary to use equations (3), (6), and (8) and carry out the indicated matrix algebra operations to obtain the variances of the estimated positions. We consider first the case of equally weighted data.

At this point in texts on statistics it is noted that the measurement noise variance $\sigma^2$ is usually unknown, so that an estimate of it must be made from the measurement residuals along with the unknown parameters. In many, perhaps even most applications, this is the best that can be achieved. The limitation of this approach is that it depends critically on whether or not the regression model used for the problem accurately models the real (physical) process. To the extent that the model is flawed, the measurement residuals will contain both measurement noise and contributions from errors in the model. But in the case of shoreline positions, we have estimates of measurement error (Crowell et al., 1991), and information about the response of beach width to storms. This knowledge can be used to analyze the data, and evaluate the model accuracy.

**ILLUSTRATIVE EXAMPLE**

To illustrate the issues discussed above, consider the sequence of shoreline positions in Table 1 taken from a typical transect at Cotton Patch Hill, DE. The RMS residual to a linear regression is 35.4 meters, more than 3 times the estimated uncertainty of the data.

Figure 1. Linear regression analysis of shoreline positions for a typical transect at Cotton Patch Hill, DE. The RMS residual to a linear regression is 35.4 meters, more than 3 times the estimated uncertainty of the data.

$y = -0.5962x + 1081.7$
$R^2 = 0.3903$

CROWELL et al. (1991) calculated maximum expected errors in the location of the high water line as digitized from maps and aerial photography. Maximum expected errors were determined by estimating worst-case error estimates for each step in shoreline location data compilation (i.e., survey, digitization, mapping, and photographic distortion errors). These errors were then combined using the theory of propagation of errors to obtain estimates of the maximum expected error in the digitized location of the interpreted shoreline (high water line). Maximum expected errors were calculated for the following data sources (all 1:10,000 scale): 1844–1930 T-sheets compiled prior to the use of aerial photography: 8.4–8.9 meters; post-1950’s T-sheets compiled using aerial photography: 6.1 meters; Non-tidal coordinated aerial photography: 7.5–7.7 meters. These figures were based primarily on analyses of T-sheets and aerial photography from Massachusetts, New Jersey, and Maryland (Calvert County), supplemented by data from New York (South Shore of Long Island) and Delaware. Importantly, short-term process variability factors, such as daily and monthly tidal changes and storm affects were not included in the error estimates.
transect at Cotton Patch Hill, Delaware. This is the same area used in DOUGLAS et al., (1998), but with two new additional shoreline positions for 1970 and 1997. This case is not an exceptional example of shoreline behavior; GALGANO (1998) and GALGANO et al. (1998) have reported a similar erosion character for beaches free of engineering changes and inlet dominance in Long Island, New Jersey, elsewhere in Delaware, Maryland, Virginia, and North Carolina.

Table 2. 95% confidence intervals (CI) for (1) all of the data and (2) 1929, 1962, and 1970 shorelines omitted. The SE of the fit is 35 M for case (1), and only 11 M for case (2).

<table>
<thead>
<tr>
<th>Date</th>
<th>95% CI (1)</th>
<th>95% CI (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1845</td>
<td>72.27</td>
<td>26.65</td>
</tr>
<tr>
<td>1944</td>
<td>26.86</td>
<td>11.12</td>
</tr>
<tr>
<td>1954</td>
<td>25.86</td>
<td>10.77</td>
</tr>
<tr>
<td>1977</td>
<td>26.78</td>
<td>11.58</td>
</tr>
<tr>
<td>1990</td>
<td>33.04</td>
<td>12.91</td>
</tr>
<tr>
<td>1993</td>
<td>34.20</td>
<td>13.28</td>
</tr>
<tr>
<td>1997</td>
<td>35.84</td>
<td>13.81</td>
</tr>
<tr>
<td>2010</td>
<td>41.74</td>
<td>15.73</td>
</tr>
<tr>
<td>2020</td>
<td>46.63</td>
<td>17.39</td>
</tr>
<tr>
<td>2030</td>
<td>51.80</td>
<td>19.14</td>
</tr>
<tr>
<td>2040</td>
<td>57.14</td>
<td>20.98</td>
</tr>
<tr>
<td>2050</td>
<td>62.62</td>
<td>22.87</td>
</tr>
<tr>
<td>2060</td>
<td>68.20</td>
<td>24.81</td>
</tr>
</tbody>
</table>

Figure 2. Propagated 95% confidence intervals (CI) for the linear regression shown in Figure 1. For this analysis, the error of the measurements was taken to be the value obtained from the residuals of the fit, 35.4 meters. The errors are so large that the 95% confidence interval would be wider than the beach a few decades after the last (1997) measurement.

There are 10 shoreline positions available at this site during the interval 1845–1997. The measurement type and estimated uncertainty are given in Table 1, as are the measurement residuals from the ordinary linear regression presented in Figure 1. The results in Table 1 are internally consistent. For example, the residual for the 1962 measurement is about 81 m, a little more than twice the ≈35 m standard deviation (SD) of the residuals. In a sample of 10 points with these statistics, the probability of occurrence of such a large residual is nearly 50%. The difficulty comes in trying to reconcile these residuals with our a priori knowledge of measurement error, seen above to be about 7–9 meters for T-sheets and aerial photography. There are, however, additional considerations.

The uncertainty of shoreline position data depends on the accuracy and precision of the survey measurements, and on the stability of the shoreline position indicator. The position indicator in the present example, the high water line, is affected considerably by a seasonal component (summer/winter), the tidal cycle, and normal day-to-day variability (SHALOWITZ, 1964; STAFFORD, 1971; DOLAN et al., 1980; PAJAK, 1997; MORTON, 1998). As an example of the latter factor, PAJAK (1997) showed that for Duck North Carolina, the standard deviation in the location of the HWL within the summer months of 1994, 1995, and 1996 was 4.1 m, 8.3 m, and 7.0 m,
respectively. The higher value (8.3 meters) occurred during the summer months of 1995, a period which saw a number of storms, including Hurricane Felix, and in which wave energy was the highest in 50 years. The average standard deviation in the location of the summer HWL for the three years is 6.5 meters, which given the extreme wave energy that occurred in the summer of 1995, is probably a conservative value—at least for Duck North Carolina.

By using summer shoreline positions and avoiding times of spring and neap tides, the high water line has proven to be a useful indicator of shoreline position on most of the US East Coast (Stafford, 1971; Leatherman, 1983; Anders and Byrnes, 1991). To be conservative, we assume that properly selected (i.e., summer beach, no extreme tides) high water line positions have a random component of uncertainty of 6.5 meters (±20% of beach width) in addition to the measurement error, making the total uncertainty of the measurements about \( (8.9^2 + 6.5^2)^{1/2} = 11.3 \) m. Thus a plausible estimate for overall measurement error is ±11 meters (less for data taken after 1977). This means that the 35 m fit to the data in the linear regression in Figure 1 is more than 3 times a conservative estimate of the uncertainty of the data. The residuals in Table 1 either represent a very improbable sample of points from a population having a standard deviation of about 11 m, or they are reflecting inadequacies of the linear trend model. Of course the latter is the case.

In the absence of a priori information concerning the model or measurement uncertainty, the best that can be done with these data is shown in Figure 2. The figure presents the trend line extended to 2060, and the 95% confidence interval (CI) about that line computed from equation (7) assuming that the variance of the measurements is the value (35 m)² obtained from the residuals. Also shown (above the zero line) is the 95% CI value by itself. Note that the 95% CI reaches nearly 70 m by 2060, a forecast of a little more than 60 years. Case (1) in Table 2 summarizes these results. For this case of 10 observations (8 degrees of freedom), the 95% CI reaches ±68

Table 3. Linear regression results without post-storm shoreline position data.

<table>
<thead>
<tr>
<th>Date</th>
<th>Position (m)</th>
<th>Residual (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1845</td>
<td>0.00</td>
<td>2.49</td>
</tr>
<tr>
<td>1944</td>
<td>-67.78</td>
<td>-7.59</td>
</tr>
<tr>
<td>1954</td>
<td>-63.66</td>
<td>2.36</td>
</tr>
<tr>
<td>1977</td>
<td>-86.92</td>
<td>-7.48</td>
</tr>
<tr>
<td>1990</td>
<td>-69.59</td>
<td>17.42</td>
</tr>
<tr>
<td>1993</td>
<td>-101.22</td>
<td>-12.46</td>
</tr>
<tr>
<td>1997</td>
<td>-85.84</td>
<td>5.25</td>
</tr>
</tbody>
</table>

SE = 11.06 M
Figure 4. 95% confidence intervals for solutions with and without the post storm shoreline data of 1929, 1962, and 1970.

Figure 5. 95% confidence intervals for the case without post storm shorelines, and with post-1977 shoreline positions taken to have uncertainties given in Table 1. Note the decrease in the values for the confidence intervals after the time of the more accurate measurements.
m (±2.306σ). For a more moderate forecast, say to 2030, the 95% confidence interval is still ±51 m. Such large confidence intervals are not very useful, since they are much larger than the erosion predicted to occur.

It is possible to obtain more useful estimates of shoreline position uncertainty due to long-term erosion than given by the uncritical application of linear regression done above. It is known that great storms occurred in 1929 and 1962 that greatly affected shoreline position. Figure 1 shows that accretion occurred for some period after the 1929 storm. In the case of the 1962 Ash Wednesday storm, accretion continued for at least a decade, as shown by the 1970 and 1977 positions. The effect on the confidence intervals of deleting the 1929, 1962, and 1970 shorelines is shown in Figure 3, and the results are tabulated in Table 3. The most obvious effect of eliminating these post-storm shorelines is that the fit to the data has been reduced to 11 m, just about the estimated error of most of the data. Since extreme data values have been eliminated, the r² value has improved from 0.39 to 0.90. Table 2 (case 2) shows that the 95% confidence intervals have also been drastically reduced (by nearly a factor of 3), even though the number of degrees of freedom has been diminished by 3. Figure 4 shows together the 95% confidence intervals for these two cases.

It is also known that very large storms occurred in 1991 and 1992 (Halloween storm). The 1993 shoreline position is landward of the historical trend, but the 1997 position is seaward. Does this justify removal of additional measurements? The answer is probably no, since the number of data points is so small and the RMS fit to the model so near the known overall uncertainty of the data. This reasoning, however, provides an a posteriori justification. What is actually required is an a priori method of selecting shorelines, perhaps one based on a measure of storm intensity.

What can be concluded from this analysis (and the much more extensive one by GALGANO et al., 1998) is that shoreline positions affected by great storms are very inconsistent with a linear trend model of shoreline retreat for an extended time that can reach even 10 years or more. But when recovery of beach width after storms can be demonstrated, as in the case here, eliminating post-storm shoreline position data gives a residual standard deviation from the linear regression that is much closer to the a priori estimates of the total measurement error. In other words, the linear trend model may hold over the long term in many cases (such as this one), even though it is seriously violated in the short-run following a great storm.

Finally, it is interesting to see the effect of the increased accuracy and precision of the measurements after 1977 on the estimated confidence intervals. As noted earlier, the general solution to the linear regression problem given by Equations (3), (6), and (8) enables treatment of data of differing weight. Figure 5 displays the results for the 95% confidence intervals for the case lacking post-storm shorelines. The data error was taken to be the values given in Table 1. As would be anticipated, there is a significant decrease in the magnitude of the 95% confidence intervals in the period subsequent to the more accurate data. In principle, it is possible to reduce the survey measurement error to the cm level or better with advanced GPS methods, but the inherent variability of the shoreline position indicator will remain as a limitation on the final accuracy of the data.

CONCLUSIONS

The most dramatic erosion events on beaches not influenced by inlets or engineering changes are those associated with great storms. In a few hours or days, the beach may erode by an amount exceeding that caused by the underlying long-term rate during a half-century. Even the normal winter/summer variation of beach width is very large compared to the secular loss of beach width over a few decades. These facts tend to obscure the inexorable nature of long-term erosion. In this paper we have presented a methodology for computing the uncertainty (confidence intervals) of predicted shoreline positions due to a long-term trend of erosion, a model appropriate to a significant fraction of the US Atlantic coastline. The scheme was easily implemented on Microsoft Excel®, which has matrix algebra capability.

The ability to compute the uncertainty of shoreline position due to long-term erosion trends is an essential element of an erosion hazard mapping program. Separating erosion/recovery events due to great storms from an underlying long-term trend of erosion leads to reasonable and possibly useful estimates of the uncertainty of long-term forecasts of shoreline position resulting from an underlying trend of erosion. In principle it is possible to modify the regression algorithm to account for the loss/recovery cycle, but it would involve many parameters, and there are simply not enough data available to estimate them. As a practical matter, valuable long-term erosion trend forecasts may be obtained in many cases from linear regressions by eliminating post-storm shoreline positions associated with great storms. The post-storm data points can then be used to determine additional erosion buffer areas or zones of excursion.

ACKNOWLEDGMENTS

We are grateful to Drs. Robert Morton and Dennis Milbert for their detailed and valuable critiques of the paper. This project was supported in part by grants from The Andrew W. Mellon Foundation and the National Aeronautics and Space Administration. Mention of a commercial product does not constitute an endorsement by FEMA, NASA, or The Andrew W. Mellon Foundation.

LITERATURE CITED


DOLAN, R.; HAYDEN, B.P.; MAY, P., and MAY, S., 1980. The reli-
ability of shoreline change measurements from aerial photographs, *Shore and Beach*, 48, 22–29.


**Zhang, K.** Twentieth Century Storm Activity and Sea Level Rise Along the U.S. East Coast and their Impact on Shoreline Position. Ph.D. Dissertation, University of Maryland, Department of Geography.