Proposed Minimum Standards for Comparing
Shoreline Model Results with Data

By : Emmett R. Foster, P.E. - FDEP/OBCS
February 12, 2002

INTRODUCTION:

The purpose of this paper is to present an objective method for determining whether a high level of statistical significance exists in a set of comparisons of actual measurements with model results. Such comparisons are used to calibrate a model and verify that it is indeed simulating a physical process such as shoreline change to a reliable extent.

This paper concentrates on comparing shoreline mean high water locations as measured by survey versus model predictions. The shoreline is probably the most important factor in public perception of success or failure, as well as being the most reliable survey data parameter. However, the procedures can also likely be adapted to compare other parameters such as volumetric change.

All of the statistical formulas suggested for use here can be readily found in most statistics textbooks and are not necessary for the intent of this paper, which is to describe the procedural logic and suggest minimum standards for passing.

Why is it important to quantify the decision of whether a model, intended to function as a calibrated model, adequately simulates shoreline changes ? The answer is that there is a perceived need to advance the discipline of shoreline modeling in this regard. There is usually a great deal of work, time, and money involved in creating a model. There are also numerous competing models and modeling techniques. If there are minimum standards to be met in demonstrating that the model works well, then the models and techniques will likely improve and adapt to produce the desired statistically significant results.

The suggested minimum performance standards are based on showing a high level of statistical significance between actual measurements and model results. The actual measurements and model results should first be highly correlated, which is a measure of how well the model simulates the overall shape or pattern of the shoreline. Second, the two should have a statistically significant one-to-one measurable correspondence. The following step by step procedures can satisfy these objectives.

PROCEDURES FOR DEMONSTRATING STATISTICAL SIGNIFICANCE:

(1) DATA SELECTION : It is suggested to use a minimum of four reliable, checked, shoreline surveys covering the modeled area. One of the four is to be the initial conditions or input shoreline. The other three are for comparison with model results. The strongest case is made by model comparison with as many reliable (per the

modeler's knowledge and judgment) data sets as are available over the relevant modeled time period. Usually the "verification" survey is one of the latest surveys. Three comparative surveys are suggested as a minimum to show the model is capable of simulating reality repeatedly. The surveys should cover a significant portion of the total time period that the physical conditions being modeled (e.g. sand supply situation, wave climate, and local geographic features such as structures) have been present. The chosen surveys should be at least one year apart and, if at all possible, several years apart.

It can be helpful, particularly when modeling over relatively short time periods of change (a few years rather than decades), to select surveys from the same season. Typically the summer season exhibits more predictable shoreline behavior because of lesser wave energy, barring a major storm event. It is suggested that immediate post storm data may not prove helpful because storm induced on-offshore changes can affect shoreline location, a process (usually) not accounted for in the physical process equations most longshore models are based on (i.e. some form of a longshore sand transport equation and conservation of volume). It is also suggested that immediate post construction surveys for nourishments are not appropriate for use as initial condition surveys because of the major on-offshore adjustments that typically occur during the first winter or high wave energy season.

(2) PLAN VIEW PLOTS : A direct visual comparison of mapped shorelines in plan view is very useful. It is necessary to plot in plan view the model generated shorelines in comparison with the measured shorelines, in the coordinate system of the model. The coordinate system of the model should of course be specified as to how it relates to a standard horizontal datum. The measured and modeled shorelines should be a close visual match. Figure 1 shows an example plan view plot.

(3) SCATTERPLOTS & CORRELATION AND REGRESSION COEFFICIENTS : Create scatterplots with equal x and y axis scales, for each of the three (or more) sets of survey versus model shoreline comparisons. In a scatterplot, the y axis (vertical axis) is for measured shoreline locations, and the x axis (horizontal axis) for model predicted shoreline locations. A calculated correlation coefficient for each plotted comparison set can be said to measure how well the model results follow the pattern of the shoreline, as seen for example in the plan views. The scatter-plotted data should appear to fall closely on or about a straight line through the data. The calculated correlation coefficient should be very close to 1.0 (greater than or equal to 0.95). Note that there can be a high amount of correlation even though there can be an order of magnitude difference between the measurements and model predictions. Therefore it is also necessary, if there is a high degree of correlation (if not, the first test is failed), to do a least squares linear fit to the data (y on x) in each scatterplot and calculate the slopes of the fitted lines. The slope is also called the regression coefficient. The slope must be very close to +1.0 (+/- 0.1) if there is a direct one-to-one comparison between measurements and model results. Note that a line with a +1.0 slope would be a 45 degree line diagonally upward on a scatterplot having axes with equal scales.

An example scatterplot showing data having suitably high correlation and regression coefficients is shown as Figure 2. An example scatterplot with a suitable correlation coefficient but unsuitable regression coefficient is shown as Figure 3. In this latter example the model follows a similar pattern as the actual shoreline but there is a 10 to 1 correspondence between the model predictions and shoreline measurements.

Note that comparing along an armored (with walls, revetments, etc.) shoreline with no beach is typically not valid; a boundary condition governs an armored shoreline and sand is not equivalent to concrete or rock. For example, sand placed in front of a wall will not likely behave as the wall did.

Also note that both the correlation and regression coefficients are dimensionless statistics. All three (or more) selected survey versus model data set comparisons should pass both the coefficient tests, and the scatterplots should clearly illustrate the results.

(4) STATISTICAL TEST OF THE DIFFERENCES : The correlation and regression coefficients are not sufficient in and of themselves to establish a high level of statistical significance. The calculated differences in shoreline location (residuals) at every comparison location for each of the three or more sets of comparisons should be statistically examined and tested. For example, if there are thirty-six DEP reference point profiles in a survey spanning the modeled area, there can be thirty-six differences calculated in mhw shoreline locations between the model predictions and the mhw survey measurements for that survey. Again, each of the three (or more) surveys must be compared and the differences collected for each set. Again note that difference measurements along an armored shoreline with no beach are typically not valid.

Create histograms (plot the number of occurrences on the y axis, within the data range blocks on the x axis, using approximately 10 to 20 data range blocks), for each set of differences. An example histogram is shown as Figure 4.

Calculate the sample mean, standard deviation, and root mean square error (a measure of the average amount of absolute error) for each of the three (or more) sets of differences. All the sample means should be tested for equivalence to zero at the 1% significance level. "Significant" means a difference unlikely to have been the result of random sampling. The following test for that assumes a gaussian distribution; this is why it is desirable to plot the histograms to see how reasonable that assumption may be. The histograms can often be inconclusive, but can also be informative.

For each set of difference data, test the hypothesis that the sample mean of the differences is equivalent to zero at the 1% significance level. This is a two-tailed "t" test for sample sizes (n) less than or equal to 31, and a "z" test for sample sizes (n) greater than 31. A dimensionless statistic using the sample mean, standard deviation, and sample size is calculated for each set of differences:

$$t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

where: $\bar{x}$= sample mean
μ = population mean (should be zero)
S = estimate of population standard deviation from sample of n observations
n = number of observations in sample

The results should be less than the critical values in the "t" table found in any basic statistics text, for the two-tailed test, (n-1) degrees of freedom, for alpha=.01 or 1% . This is the minimum (easiest to pass) test for statistical significance. All three (or more) difference sets should pass the test.

Ideally we want all the sample means (averages) of the differences between the actual measured shoreline locations and the model predictions to effectively be zero. In statistical terms, the “null hypothesis” is that the apparent difference between the sample mean and the hypothetical population mean of zero is simply the result of sampling error, given a relatively small sample size. If the test is failed, the null hypothesis is rejected and the result is said to be significant at the 1% level. This indicates we would wrongly reject the null hypothesis only 1 out of 100 times, which is a considerable probabilistic argument against the null hypothesis. This would indicate the sample is affected by something other than sampling error. This can occur, for example, when trying to match up with a post storm survey and finding that the shorelines are offset from each other by some average amount not reasonably equivalent to zero. Alternatively, the model may be producing an offset or otherwise poorly matched shoreline.

If one or more tests fail, it is possibly helpful to see if the results can be significantly improved by removing a small number of outlying difference data from one small zone along the shoreline. If so, show the change in statistical results and consider the practical significance of this finding. The significance might be the reduced or questionable reliability of the model in a certain location, such as next to a difficult boundary condition, or a suspicion of lesser quality data in a particular zone in a particular data set. Either of these should be noticeable in the plan view shoreline plot comparisons.

Review the magnitudes of the sample standard deviations and rms error values. They should be reasonable in magnitude relative to observed shoreline conditions in the modeled area and/or the types of survey data used.

(5) REPORTING RESULTS : Provide all results in a report including the name of the modeler in responsible charge. Include all the plots and clearly state the results of the tests. If at least three selected comparison sets pass the tests, the model results may be considered to be statistically significant and acceptable. Clearly state any limitations deriving from the results which were necessary to achieve the minimum standards and which affect the applicability of the model. State any expected limitations on the applicability of the model for any reason.

SUMMARY:

It is important for the advancement of the shoreline modeling discipline to demonstrate statistically significant comparisons between actual measurements and model predictions, to objectively determine if a "calibrated / verified" model effectively simulates reality. This can be done by careful selection of the data to be compared, by testing the correlation and regression coefficients of the compared data, and by performing a means test on the differences between the actual measurements and model results on at least three sets of comparisons. These tests are accompanied by graphical representations of the data : plan view plots, scatterplots, and histograms. If all tests are passed and any known model conditional limitations defined, the model will have been shown to have an acceptably high level of statistical reliability in predicting shoreline changes. The conditional limitations should be carefully reviewed to be assured the model remains useful for the intended applications. The procedures can likely be adapted to test other modeled parameters, if desired.

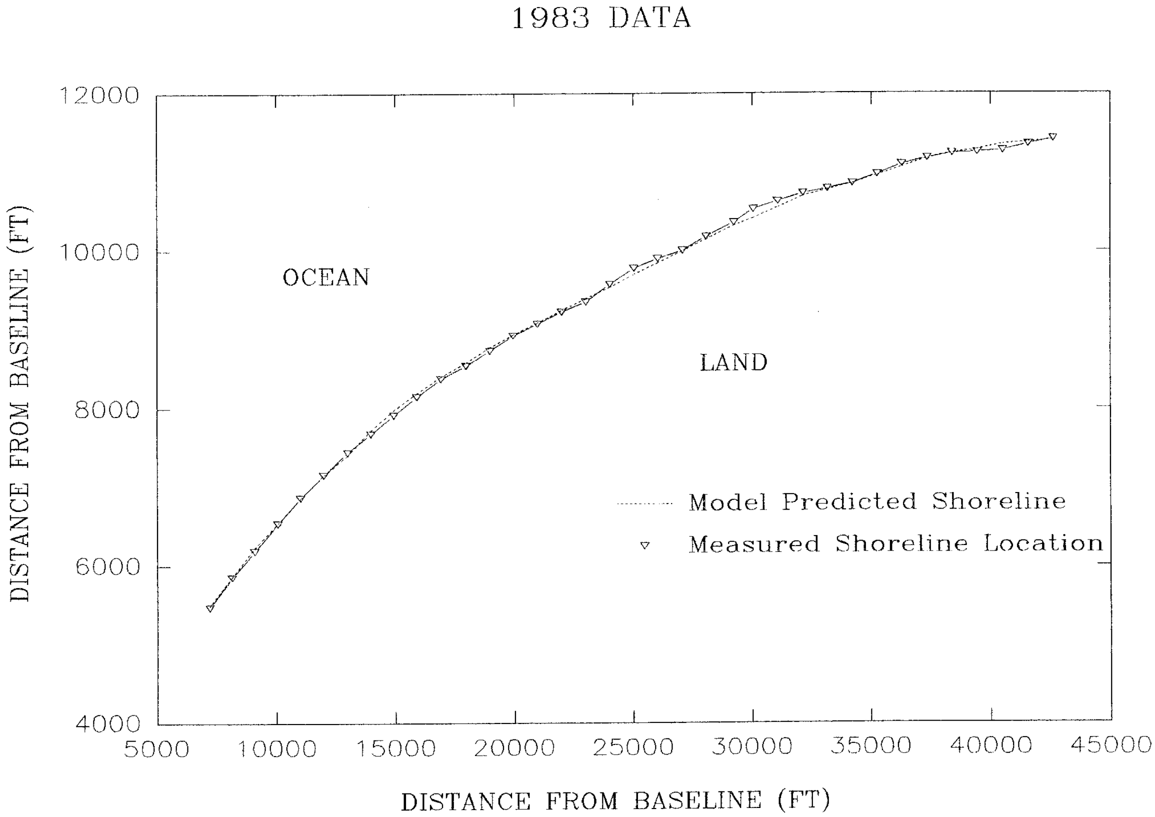

FIGURE 1 – EXAMPLE PLAN VIEW
Baseline Origin : N 242213.047 (st.pl. 27)
E 1731132.143
Bearing : 10.5 Deg. CCW from N

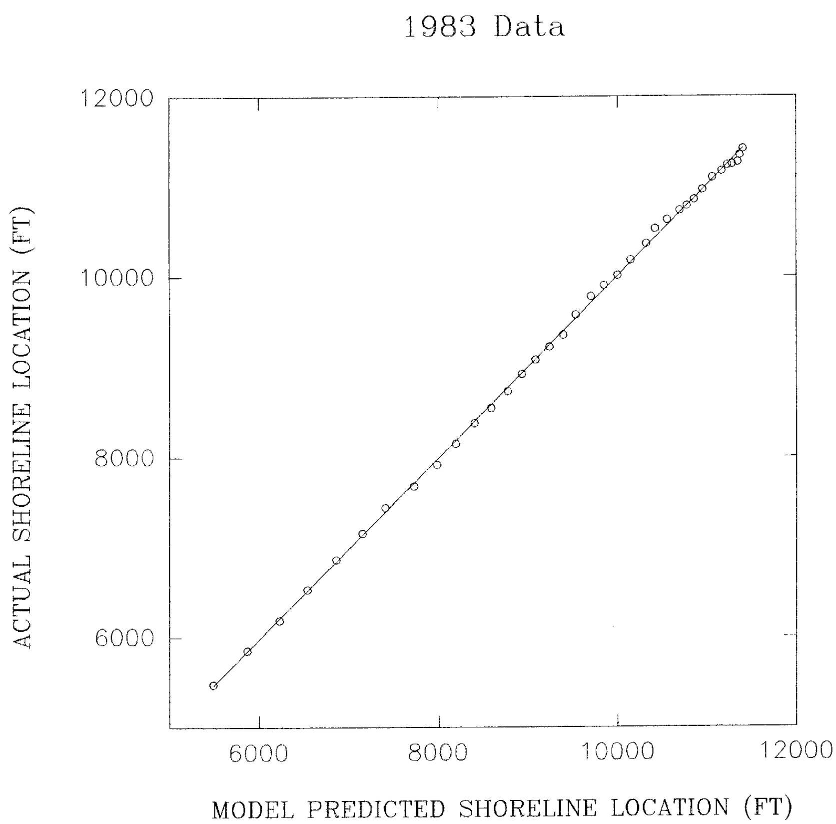

FIGURE 2 – EXAMPLE SCATTERPLOT
Regression Coef. (Slope) = 1.0071
Correlation Coef. = .9997

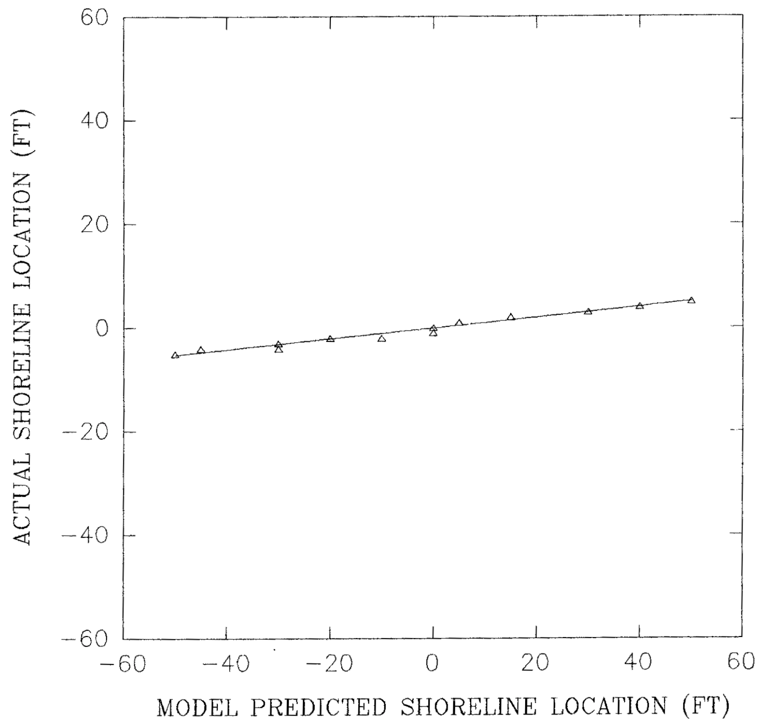

FIGURE 3 – EXAMPLE SCATTERPLOT
Regression Coef. (Slope) = 0.103 = Fail
Correlation Coef. = .987

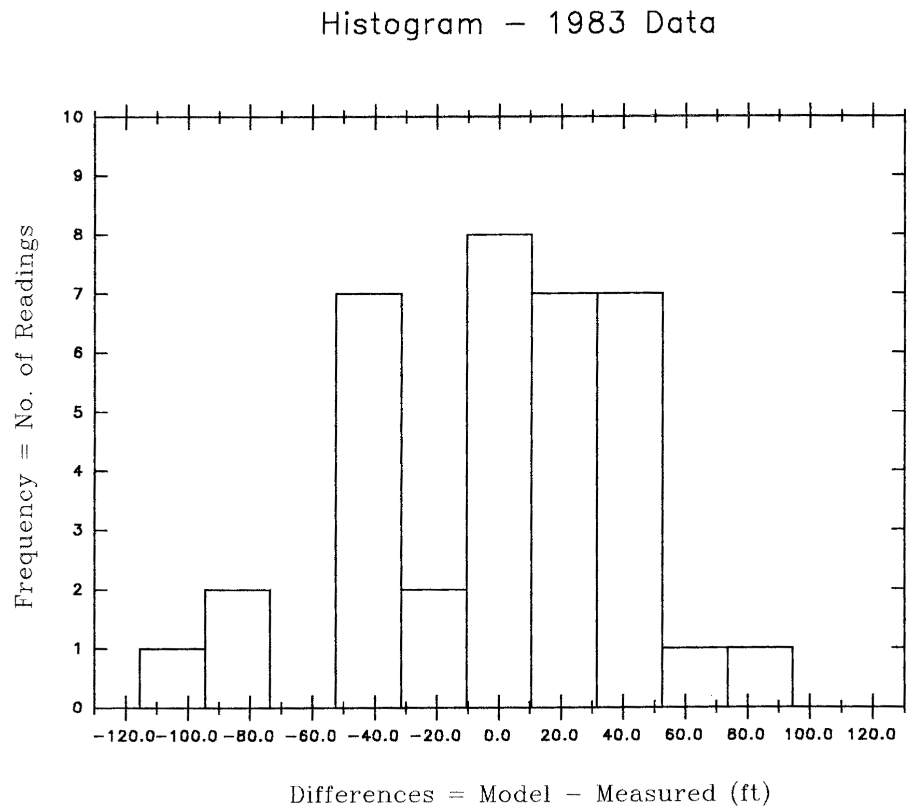

FIGURE 4 – EXAMPLE HISTOGRAM
Sample Size = 36
Mean = –0.3 ft
Std. = 41.5 ft
z = .0434 << 2.58 Critical Value (0.01)