METHOD FOR ANALYZING OBSERVED DATA in TIDAL WATERS

INTERSTATE SANITATION COMMISSION

New York

New Jersey

Connecticut

METHOD FOR ANALYZING OBSERVED DATA in TIDAL WATERS

*Excerpted from
1958 Annual Report
of the
Interstate Sanitation Commission

METHOD FOR ANALYZING OBSERVED DATA in TIDAL WATERS

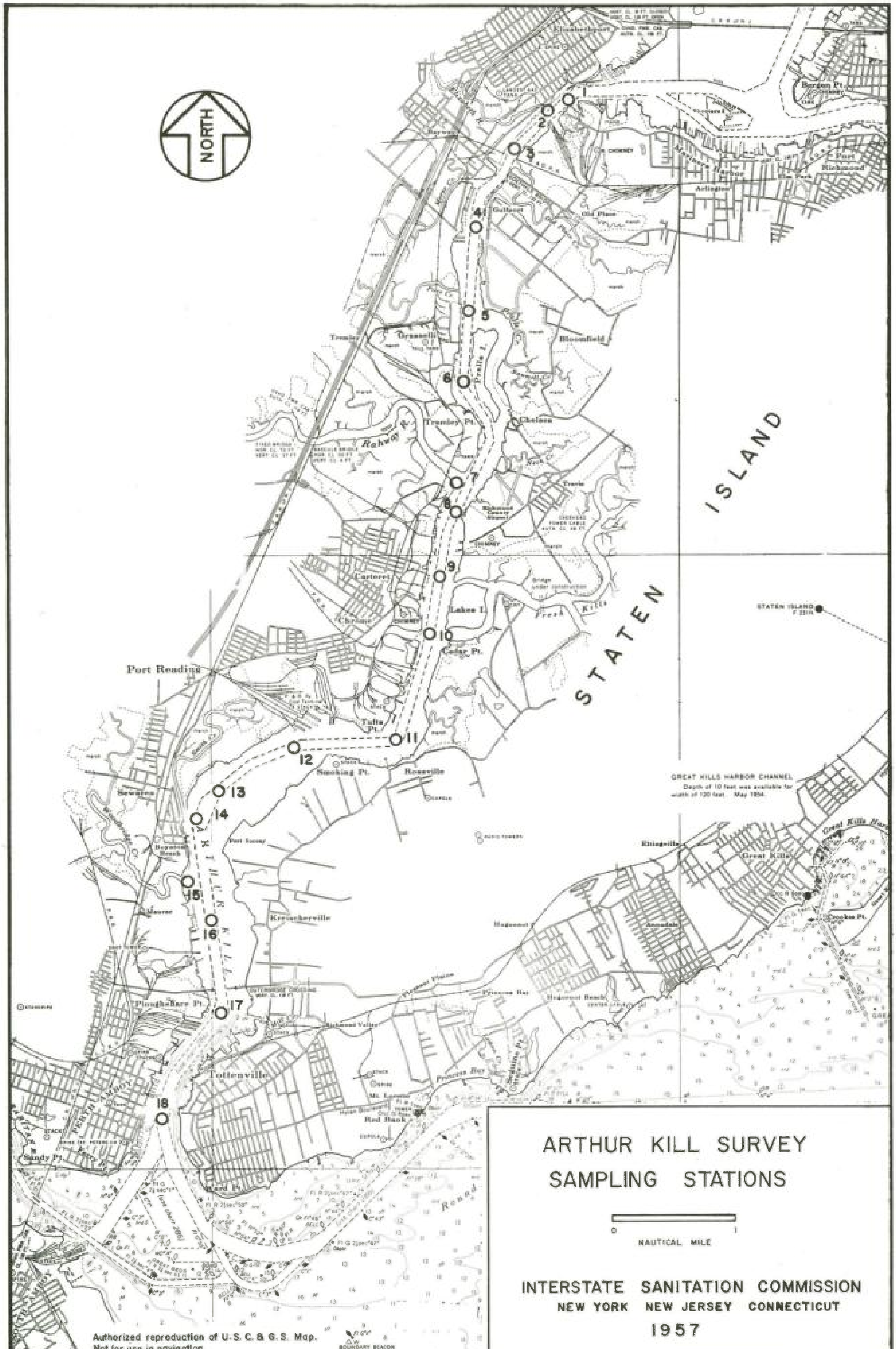
INTRODUCTION

The Interstate Sanitation District is composed of certain tidal waters common to and affecting the States of New Jersey, New York and Connecticut. In fulfilling its obligations to the participating states, it has felt the need of a quantitative method of analysis for determining the condition of these waters. This Commission and many other groups have observed the variations in dissolved oxygen produced by tidal currents. This year a study of data obtained from a water survey of the Arthur Kill resulted in the derivation of such a method of analysis which will be presented here.

SURVEY PROCEDURE

The Compact of the Interstate Sanitation Commission requires that in its Class "A" waters an average dissolved oxygen content be maintained at a depth of about five feet below the surface, of not less than fifty percent (50%) saturation during any week of the year and in Class "B" waters not less than thirty percent (30%) saturation during any week of the year.

In planning a survey to determine the compliance of any section of tidal waters, it was found necessary to locate sampling stations at more or less regular distances along the main profile or profiles through the waters to be surveyed. If these waters are narrow and have a dredged channel running through them, then the survey stations should be located in the middle of the channel. It was found that stations in the middle of the channels are sensitive and reflect any changes in water quality produced by tidal actions. When additional information is required along the main profile, close to shorelines, in tributary streams or rivers and at outfalls, additional sampling stations can be added to supplement this main profile. The map, Arthur Kill Survey Sampling Stations,



ARTHUR KILL SURVEY
SAMPLING STATIONS

0 NAUTICAL MILE 1

INTERSTATE SANITATION COMMISSION
NEW YORK NEW JERSEY CONNECTICUT
1957

Authorized reproduction of U.S.C.&G.S. Map.
Not for use in navigation.

shows how the sampling stations were spaced for a particular survey. It may be necessary to use some other distribution of sampling points in large water areas such as large bays where the influence of ship channels is not too significant. In this latter case, particular care should be taken to locate stations in such a manner that all samples will be taken at the same point each time. This means selecting good landmarks, range lights and marker buoys for good alignments. The maximum distance allowable between any two stations will, of course, depend upon the thoroughness desired in the waters being surveyed.

In sampling a station, the boat was required to stop and cut its propellor to eliminate any agitation of the waters. After all signs of agitation had disappeared from the immediate waters, the sample was taken with a Foerst Water Sampler which has a capacity of 2.2 liters. This sampler consists, essentially, of a cylindrical brass container which has facilities for opening and closing both ends. It is raised or lowered by means of a heavy cord which is graduated in five foot intervals. In taking a water sample at a predetermined depth, the operator first sets the sampler in the open position. He holds the messenger in one hand and puts the sampler over the side of the boat. It is then lowered by means of the graduated cord to a predetermined depth. Once the sampler is at the correct depth, the operator releases the heavy messenger which slides down along the cord until it hits the closing mechanism. This causes the rubber stoppers to close both ends of the sampler. This traps the water at a particular depth in the cylinder and the operator then retrieves the sampler by means of the cord.

Once at the surface, the sample is discharged into standard dissolved oxygen bottles by means of a brass tube with extension tube and release valves. This enables the operator to withdraw the water as desired and without aeration. About 900 milliliters of a sample should be allowed to flow through each dissolved oxygen bottle. A small amount of each sample is retained for analyses in the Commission laboratory. Temperature of a sample is determined immediately by means of a Weston Dial Thermometer. All dissolved oxygen determinations are made immediately on board the boats while pH and chlorides are determined at the Commission laboratory.

The Commission in two different surveys arranged to sample each station three times a day within a six hour interval and over a period of several weeks. For best results, water pollution surveys in tidal waters require that a large number of samples be taken.

ANALYSIS OF DATA FOR A SINGLE STATION

General

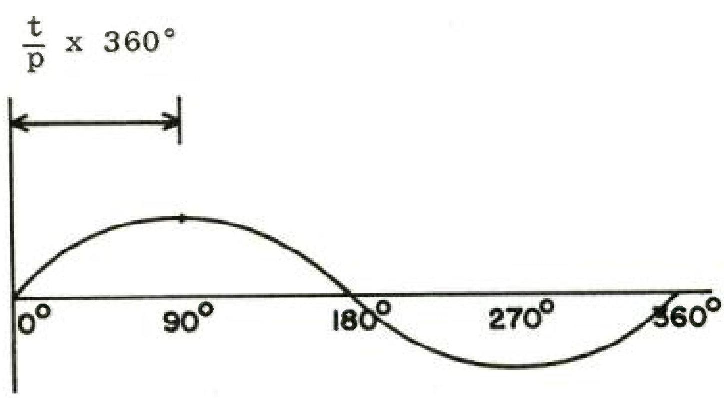
The Compact requirements of the Interstate Sanitation Commission require that certain minimum percentages of saturation of dissolved oxygen be maintained in its various waters. This mode of expression is desirable for direct interpretation for compliance with Compact standards and for purposes of quantitative analysis in that it considers the variables of temperature, chlorides and dissolved oxygen. Thus, a percent saturation of dissolved oxygen is calculated for each sample taken.

The tidal current cycle has a sinusoidal variation and the percent saturation of dissolved oxygen values have been observed to vary, in general, in a sinusoidal manner relative to it. In a program of randomized sampling, samples are taken over a period of days and from many tidal current cycles of different lengths. This type of sampling is considered advantageous in that you are not attempting to sample one specific part of a current cycle. By sampling a given station at approximately the same hours each day you gradually sample all phases of a current cycle, as the time of maximum ebb current changes continuously. Thus, the variations in percent saturation existing during all stages of the cycle will be observed in time.

Reduction of Observations
to One Theoretical Sine Curve

To reduce the observations from various tidal cycles to one common theoretical cycle, which has the distribution of a sine curve, it is necessary to refer all observations to the same current condition. The mean percent saturation of dissolved oxygen, under these conditions, occurs at the time of maximum ebb and flood currents. Here the maximum ebb current is used as the reference time. Thus, to reduce the observations in various tidal cycles to one common theoretical cycle, which has the distribution of a sine curve, you determine the lapse in time, i.e., hours between the time of observation and the time of the last maximum ebb current. This time is referred to as "t". The length of the particular current cycle is obtained from the "Current Tables for the Atlantic Coast of North America," and is referred to here as "p". To place a particular observation on the sine curve, we use the expression θ, in degrees, $= \frac{t}{p} \times 360°$.

This is shown graphically by the following example:



Where:

(1) Observation is made at 0900 hours EST.

(2) Time of last maximum ebb was at 0600 hours EST.

(3) Length of the current cycle in which sample is taken is equal to p = 12.0 hours.

(4) The difference in time between the time of observation and the time of the last maximum ebb current is referred to as "t". "t" = 0900 -0600 = 3 hours 0 minutes or expressed decimally "t" = 3.00 hours.

(5) The position of the curve then becomes:

$$\frac{t}{p} \times 360° = \frac{3.00}{12.00} \times 360° = 90°$$

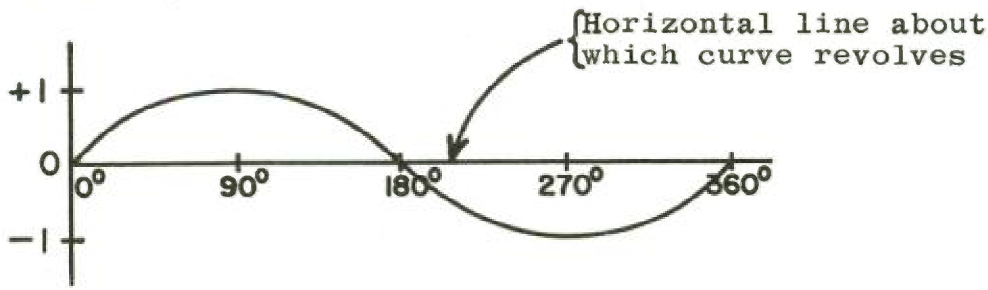
(as shown on preceding sketch)

In a similar manner, the foregoing procedure is used to obtain a distribution curve for each station which expresses all observations on a common and comparable basis.

A graph of the observed data should be plotted showing Percent Saturation (of dissolved oxygen) versus Degrees on the horizontal scale. The variation of the observed percent saturation as shown by such a graph for a particular station and as considered thus far is best described by a sinusoidal curve of best fit. However, certain geographical conditions, tidal actions and/or concentrations of pollution may cause the distribution of observed data for a few stations to be best described by a straight line, parabola, or various combinations of both. These curves or lines are determined by the method of least squares and regression analyses are made to determine the significance of these lines of best fit. The principal curve with which we are concerned is the sine curve.

Derivation of Equations

The general equation for the sine curve may be expressed as: $Y = \sin \theta$. The theoretical curve starts at 0 and increases to a maximum of +1 and decreases to a minimum of -1, in revolving about a horizontal or curve mean in passing from 0° to 360° (or 0 to 2π radians). Graphically this would appear as follows:



To adopt this general equation to our needs it will be necessary to do the following:

(1) Place the horizontal line about which the sine curve revolves through our data and at that level which represents the weighted mean of our observed data. This line may then be called "K_1".

(2) Determine the amplitude of the data so that the curve in revolving about the mean will have the correct range or variability. This amplitude may be called "K_2".

The general equation applicable to our needs may then be expressed as:

$$Y = K_1 + K_2 \sin \theta$$

Where:

Y = the observed percent saturation of dissolved oxygen.

K_1 = weighted mean percent saturation of dissolved oxygen.

K_2 = amplitude of the sine curve of best fit for the observed data.

$$\theta = \frac{t}{p} \times 360°$$

t = the elapsed time (hours) between the time of observation and the time of the last maximum ebb current.

p = length of the respective current cycle (hours) during which the particular observation was made.

In its publication, "Current Tables for the Atlantic Coast of North America," the U.S. Department of Commerce, Coast and Geodetic Survey gives the various times of maximum ebb currents at various points. Using these times it is possible to determine the respective values of "t" and "p" as defined above.

Using the method of "least squares", the weighted mean (K_1) and amplitude (K_2) of a curve of best fit for the observed data can be determined. We shall start with our general equation expressed as follows:

$$(1) \quad Y = K_1 + K_2 \sin \theta$$

Then, for a particular sample, the observed percent saturation of dissolved oxygen, Y, is equal to the predicated value, $K_1 + K_2 \sin 0$, plus some error " ".

Thus, our general equation (1) becomes:

$$(2) \quad Y = K_1 + K_2 \sin \theta + \epsilon$$

and

$$(3) \quad Y - K_1 - K_2 \sin \theta = \epsilon$$

Using the method of "least squares" and summing up the errors, we have:

$$(4) \quad \sum(Y - K_1 - K_2 \sin \theta)^2 = \sum \epsilon^2$$

Taking the partial derivative of the latter function with respect to the constants K_1 and K_2 we get the two normal equations to be:

$$(5) \quad K_1 n + K_2 \sum \sin \theta = \sum Y$$

$$(6) \quad K_1 \sum \sin \theta + K_2 \sum \sin^2 \theta = \sum Y \sin \theta$$

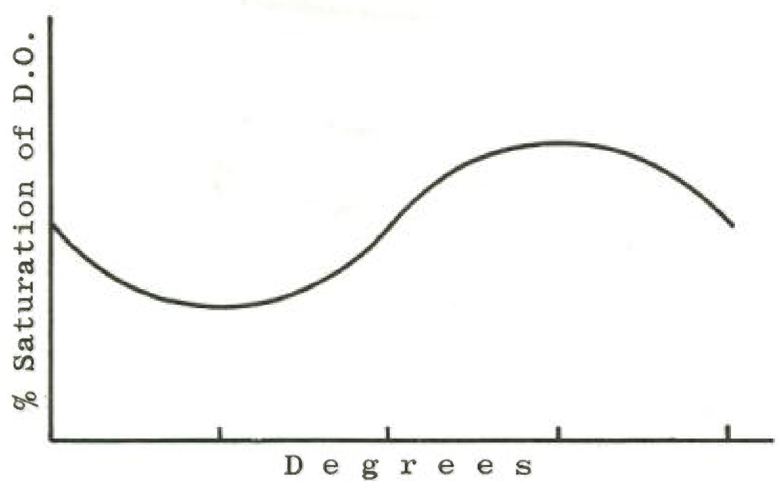
Using the method of determinants, the values of the above constants, K_1 and K_2 may be expressed as follows:

$$(7) \quad K_1 = \frac{(\sum Y)\ (\sum \sin^2 \theta) - (\sum \sin \theta)\ (\sum Y \sin \theta)}{(n)\ (\sum \sin^2 \theta) - (\sum \sin \theta)\ (\sum \sin \theta)}$$

$$(8) \quad K_2 = \frac{(n)\ (\sum Y \sin \theta) - (\sum Y)\ (\sum \sin \theta)}{(n)\ (\sum \sin^2 \theta) - (\sum \sin \theta)\ (\sum \sin \theta)}$$

These latter values of K_1 and K_2 may be used on all stations where the distribution of data is in phase with the sine curve.

Observed data for a particular station has also been observed to distribute itself as appears below:



The latter figure demonstrates how the results were 180° out of phase with the previous general equation. Thus, in rewriting our general equation to fit this condition, we have:

$$(9) \quad Y = K_1 + K_2 \sin(\theta + 180°)$$

Following the same procedure as previously, new expressions for the values of the constants K_1 and K_2 can be found to be:

$$(10) \quad K_1 = \frac{-(\sum Y)(\sum \sin^2 \theta) + (\sum \sin \theta)(\sum Y \sin \theta)}{-(n)(\sum \sin^2 \theta) + (\sum \sin \theta)(\sum \sin \theta)}$$

$$(11) \quad K_2 = \frac{(n)(\sum Y \sin \theta) - (\sum Y)(\sum \sin \theta)}{-(n)(\sum \sin^2 \theta) + (\sum \sin \theta)(\sum \sin \theta)}$$

Once the values of K_1 and K_2 have been determined, substitute them into the equation

$$Y = K_1 + K_2 \sin(0 + 180°)$$

This latter is then the equation of the curve of best fit for a particular station. When the observed data is distributed in such a manner as to be best described by one or more straight lines, the data may still be expressed in degrees but referred to as "X" in determining the line of best fit.

A straight line of best fit may be applied to the observed data using the following general equation:

$$(12) \quad Y = b_o + b_1 X$$

Where:

Y = observed percent saturation

b_o = constant

b_1 = slope of line of best fit

X = independent variable expressed in degrees

Using the method of "least squares" and taking the partial derivatives of the function with respect to b_o and b_1 we get the two normal equations:

$$(13) \quad nb_o + b_1\sum X = \sum Y$$

$$(14) \quad b_o \sum X + b_1\sum X^2 = \sum XY$$

Here, the values of n, $\sum X$, $\sum Y$, $\sum X^2$ and $\sum XY$ are obtained from the observed data. However the two normal equations are solved for the constants b_o and b_1 using the matrix method. A regression analysis is made to determine the sums of squares and mean squares assocated with b_o and b_1. The "F" test is then applied to determine the significance of b_o and b_1.

Correction of Amplitude

In the previously mentioned publication, "Current Tables for the Atlantic Coast of North America", the times of occurrence of maximum ebb currents, at various reference points, are listed for each day of the year. Thus, it is possible to estimate a time correction or time difference which can be applied to one of these reference points to obtain the times at which the corresponding maximum ebb currents will occur at a particular sampling station. Then, with the times at which all samples were taken and the times of maximum ebb currents, the results are put on a comparable basis and in this case plotted on a graph showing Percent Saturation vs.Degrees (the latter expresses the proportion of current cycle completed.) Once this graph is completed, it is possible to use it in determining how accurate the estimated time difference was for a particular station. This time correction is usually different for each station, provided that the stations are not located close to one another.

Once the data has been plotted on a graph as described and a sinusoidal curve of best fit has been determined for the data, then the calculated curve of best fit with mean should be plotted and drawn on a piece of tracing paper. This is superimposed on the graph of observed data. Keeping the line representing the calculated mean directly on the line of the graph corresponding to this value, more the tracing paper horizontally to the left or right until the curve of "best fit" is in phase with the observed points. Then, read directly the number of degrees between the zero point for the graph of observed data and the zero point for the curve of best fit (on tracing paper). This is a graphical determination of the error (in degrees) of the time of occurrence of maximum ebb or the degrees out of phase.

Once a graphical solution is made of the error or number of degrees that the sine curve is out of phase, the calculated amplitude may then be corrected. This may be done by using the following expression:

$$K2_{corrected} = \frac{K2_{calculated}}{\sin\,(90 - \theta_1)}$$

Where:

$K2_{calculated}$ = the amplitude of the sine curve of best fit calculated when the time of occurrence of maximum ebb current is in error.

θ_1 = error (in degrees) in the time of occurrence of maximum ebb current or degrees out of phase with either of the two conditions considered.

$K2_{corrected}$ = the amplitude of curve of best fit when time of occurrence of maximum ebb current is correct.

Effect of Time Corrections on Mean

Once an estimate has been made of the time of occurrence of maximum ebb current at a sampling station, the mean of the data can be computed. If the estimation is in error, it will not significantly affect the mean value. This is better realized if you consider that for any given sinusoidal curve or set of data, it revolves about the same mean irrespective of the out of phase angle.

Determination of Simultaneous Conditions at Several Stations

Once the corrected curve of best fit has been determined for the stations, each is plotted on graph paper. Then, select any particular twelve hour period on any day during the time of survey. On the day selected, determine the time of maximum ebb current for one of the main sampling stations and let this be the hour or time at which the twelve hour period starts. This time of maximum ebb current then would become the time of the first observation for all of the stations under consideration. The time for the next observation will be one hour later. The time for each of the succeeding observations will be one hour later than the preceding observation. This procedure will be continued until twelve consecutive hourly observations have been made. The time of the last maximum ebb current for the day under consideration can be computed for each of the stations.

A table with the following column headings should be set up for tabulation of data and completion of necessary computations: Sampling Station, Time of Observation (E.S.T.), Time of Last Maximum Ebb Current (E.S.T.), Time Elapsed Since Maximum Ebb Current in hours ("t"), Length of Cycle in hours ("p"), Point on Curve of Best Fit (Degrees), and Percent Saturation of Dissolved Oxygen.

This procedure in effect will produce the same results as sampling all of the stations at the same time for a given observation and thus determining the conditions at a particular time throughout the waterway.

Using the current tables and applying the time corrections determined for each station, you can compute the times at which the same maximum ebb current will occur at the respective sampling stations. This is then the time of the last maximum ebb current. Now subtract this latter time from the selected time of observation. This difference must be expressed in decimals of an hour and is referred to as "t". The length of cycle is referred to as "p" and expressed as decimals of an hour also. The latter is easily determined as explained in an earlier section. Observation time selected at the beginning and end of a current cycle may, after applying the time corrections, extend over into another current cycle. If this occurs, then it is necessary to calculate the length of the cycle in which it falls and use that value as "p" for the particular current cycle.

The next step is to determine the point on the curve of best fit which corresponds to the conditions existing at the sation at the selected time of observation. This is done using the following expression:

$$\text{Point of curve of best fit} = \frac{t}{p} \times 360°$$

This point is expressed in degrees and is for a particular time and station. Using the graph of the curve of best fit for the respective station, determine that percent saturation of dissolved oxygen corresponding to the number of degrees calculated for the point. This procedure is then repeated for each station.

A percent saturation of dissolved oxygen versus distance (in miles) profile should be made of all main sections through the waterways. Place the distance in miles on the horizontal scale and mark on it the location and number of each sampling station (for example, see Plate I). Plot the percent saturation values just determined on the profile for each station for each selected time of observation. The authors found that the selection of hourly interval times of observation within twelve consecutive hours gave good results. Since the current cycle is approximately twelve hours in duration, this allows for six different percent saturation profiles or conditions for the ebb current phase and six for the flood current. See Plates II and III as an example of the hourly variations observed in an actual survey.