A WATER-QUALITY SIMULATION MODEL FOR WELL MIXED
ESTUARIES AND COASTAL SEAS: VOL. VI, SIMULATION, OBSERVATION,
AND STATE ESTIMATION

JAN J. LEENDERTSE
SHIAO-KUNG LIU

R-1586-NYC
SEPTEMBER 1974

THE NEW YORK CITY
RAND INSTITUTE
545 MADISON AVENUE NEW YORK NEW YORK 10022 (212) 758-2244
PREFACE

The research presented in this report is part of a continuing research program in water-quality management of Jamaica Bay for the City of New York. In this volume, the scope of the investigation has widened from simulations of Jamaica Bay itself to an analysis of the whole drainage system of the bay. Observations in and around the bay were used, together with simulation of the separate drainage and sewer systems' response to rainfall; subsequently, simulations were made of the response of the bay to rain-induced inputs from the combined storm sewer overflows and other discharges. The details of the behavior of these drainage systems are presented in a separate report. It is our intention to report later on the technical alternatives for optimum control of discharges and management of the system as a whole.

This report was originally ready for publication in August 1973. Mr. Martin Lang, now First Deputy Administrator of the Environmental Protection Administration of the City of New York, suggested that publication be delayed in order to include the important results of another experiment. This experiment, intended as an additional verification of the model, was similar to that described in this report except that the observed water-quality data was retained by the City until our predictions were completed. The comparison graphs of computed and observed state estimates were prepared in New York just prior to a briefing given to the First Deputy Administrator of the Environmental Protection Administration and the Commissioner of Water Resources. This comparison in many respects exceeded the good results described in this report, and appears to warrant a separate publication. The report is now in preparation.

We have taken new approaches to the analysis of the complex problem of pollution control, often adapting or extending methods used in other fields of engineering and science. The developments and analyses we have made have often been owed to perseverance; frequently results have seemed imminent, then have eluded us until complicated mathematical and computational problems could be overcome. Nevertheless, the Jamaica Bay study gives both authors much professional satisfaction, and the results of the laborious efforts in methodology and program development will undoubtedly contribute to the solutions of pollution problems in other domestic and foreign estuaries, as is already evidenced in published applications of the model. *

During our research we have been confronted with the problem of inducing change in the methods and institutions of the tidal hydraulics engineering profession in the United States. Although resistance to such change has been considerable, it has now become clear that changes are inevitable; large-scale computer simulations and methods of extended data analysis are now being introduced into this profession.

We are much indebted to an enlightened City Administration, which recognized early that conventional methods would not lead to a solid basis for decisionmaking in New York City's costly program of water pollution control and enabled us to work on this problem of fascinating complexity.

* 東京湾の保全と開発計画に関する調査 一 野村研究研究所 一 昭和43年3月

SUMMARY

The report describes the water-quality simulation of post-rainstorm coliform bacteria distributions in Jamaica Bay, New York, by use of models of the drainage basins surrounding the bay and a water-quality simulation of the bay itself.

A stochastic analysis method was introduced into the investigation to assess the behavior and resolving power of the water-quality simulation model and to derive an optimal estimate of missing input data.

The estimates obtained by simulation agree well with those obtained by field measurements, except near Bergen Basin, where an unknown source of coliform bacteria exists. The origin and extent of this input should be determined from new field surveys.

Since the response to a rainstorm of all major components of this urban estuarine system can be determined, the models described in this report will provide the basis for the optimal design and management of an auxiliary treatment system for sewer overflows of the drainage basins around the bay.
ACKNOWLEDGMENTS

We are much indebted to Mr. Martin Lang, First Deputy Administrator of the Environmental Protection Administration in New York City, for his support and encouragement of our studies. We particularly appreciate his critical review of the manuscript of this report.

During the study period Mr. Charles Samowitz was Commissioner of Water Resources. We also want to acknowledge the cooperation of Mr. Norman Nash and Mr. William Pressman of the Department of Water Pollution Control. Through their efforts and guidance in the field investigation we obtained the valuable biological data which made this study possible. Messrs. S. Arella, M. Singer, and H. Innerfeld of this department made the field investigation.

The execution of this study would not have been possible without our Rand colleagues Mr. A. Nelson, Ms. C. Johnson, and Ms. M. Lakatos, who prepared the computer programs for the simulations, data management, and graphic displays.

Ms. J. Douglas expertly and professionally prepared the numerous drafts of this report with great patience.
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREFACE</td>
<td>iii</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vii</td>
</tr>
<tr>
<td><strong>I. INTRODUCTION</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>II. ANALYSIS IN THE FREQUENCY DOMAIN</strong></td>
<td>5</td>
</tr>
<tr>
<td>Linear Systems</td>
<td>5</td>
</tr>
<tr>
<td>Cross-Spectrum</td>
<td>10</td>
</tr>
<tr>
<td>Uncorrelated Components</td>
<td>12</td>
</tr>
<tr>
<td><strong>III. CROSS-SPECTRAL ANALYSIS OF TIDES AND TIDAL FLOW</strong></td>
<td>15</td>
</tr>
<tr>
<td>Energy Dispersion</td>
<td>15</td>
</tr>
<tr>
<td>Use of Response Function for Model Adjustment</td>
<td>20</td>
</tr>
<tr>
<td>Determining System Behavior by Spectral Methods</td>
<td>23</td>
</tr>
<tr>
<td>Energy Distributions in Higher Frequencies</td>
<td>27</td>
</tr>
<tr>
<td>Reconstruction of a Missing Record</td>
<td>31</td>
</tr>
<tr>
<td><strong>IV. STATE ESTIMATION OF COLIFORM DISTRIBUTION BY</strong></td>
<td>38</td>
</tr>
<tr>
<td>OBSERVATION AND SIMULATION</td>
<td>38</td>
</tr>
<tr>
<td>Observed Inputs and Other Parameters</td>
<td>40</td>
</tr>
<tr>
<td>Peripheral Inputs</td>
<td>45</td>
</tr>
<tr>
<td>Simulation</td>
<td>81</td>
</tr>
<tr>
<td><strong>V. DISCUSSION</strong></td>
<td>81</td>
</tr>
<tr>
<td>Simulation</td>
<td>82</td>
</tr>
<tr>
<td>Water-Quality Management and Control</td>
<td></td>
</tr>
<tr>
<td><strong>VI. CONCLUSIONS</strong></td>
<td>83</td>
</tr>
<tr>
<td><strong>Appendix</strong></td>
<td>85</td>
</tr>
<tr>
<td>A. RESPONSE FUNCTION DETERMINATION BETWEEN DENSE</td>
<td>88</td>
</tr>
<tr>
<td>AND SPARSE DATA SETS</td>
<td></td>
</tr>
<tr>
<td>B. EFFECT OF TIDAL GAUGE CLOCK ERRORS ON VERIFICATION</td>
<td>93</td>
</tr>
<tr>
<td>RESULTS</td>
<td></td>
</tr>
<tr>
<td>C. COLIFORM VARIATIONS IN TIME AT DIFFERENT LOCATIONS</td>
<td>99</td>
</tr>
<tr>
<td>IN JAMAICA BAY</td>
<td></td>
</tr>
<tr>
<td>D. PROPOSED MONITORING PROGRAM</td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>103</td>
</tr>
</tbody>
</table>
I. INTRODUCTION

Jamaica Bay is a bay with an area of about 20 square miles, consisting partly of marsh and partly of dredged channels and basins, located at the southwestern end of Long Island within the City of New York.

The bay area serves many diverse needs. Along its boundaries are residential areas and major airports. The channels through the bay are heavily traveled navigation pathways supplying local gasoline and heating oil as well as fuel to John F. Kennedy International Airport. A unique wildlife refuge is located in the center of the bay. The area has a large recreational potential, though this has been impaired by the unsatisfactory water quality of the bay.

At present a program is under way to upgrade existing water pollution control facilities and to terminate the discharge of untreated combined sewer overflows after rainstorms by constructing Auxiliary Water Pollution Control Facilities (AWPCFs).

The first such facility built is the plant for the Spring Creek Basins (see Fig. 1). In contrast to the three conventional treatment facilities (WPCFs) around the bay, which are operated continuously, these auxiliary facilities only become operative after minor rainstorms, when the trunk sewers or the WPCF cannot handle all flow. At this time the excess is stored in the AWPCF, and treatment is initiated. After not too intense rainstorms, the stored water is pumped back into the system and further treated by the conventional facility. Only after intense storms will the facility discharge into the bay after disinfection of the effluent.

For an optimum design of AWPCFs along the northern shore of the bay with respect to required storage and treatment characteristics and interconnections with other facilities, it is necessary to know the responses to rainstorms of the drainage basin in quantity and quality of overflow, as well as the water-quality effects caused by discharges into the bay. To accomplish this analysis, a water-quality simulation model of Jamaica Bay was developed, as well as models for each of the drainage systems indicated in Fig. 1. The boundaries of the drainage systems near Kennedy International Airport are not known exactly, and only the discharge points are indicated on the figure.

This report presents the observation, simulation, and state estimation of coliform distributions in Jamaica Bay resulting from a rainstorm during the period from May 31 to June 3, 1972. In this period the rainfall in the vicinity of the bay was observed, together with the coliform water quality in the bay. The effort described here is a major part of the research program. Not only are two types of models evaluated simultaneously—i.e., the water-quality simulation model for estuaries and a model for computation of overflow quantity and quality of each of the drainage basins around Jamaica Bay—but also a stochastic analysis of observation and simulation is introduced. These stochastic analysis methods, together with the deterministic water-quality simulation model, made it possible to obtain state estimates of quantifiable variables in the estuary. The use of these models in later studies will lead to new methods of water-quality management through the assessment of technical alternatives—in both a strategic and tactical sense—and to new methods of optimal control of water pollution.

The classical method for evaluating hydraulic and water-quality simulation models has been to compare field measurements with simulated results. Generally, such a comparison is made subjectively by comparing data in graphic form, then
Fig. 1—Jamaica Bay urban-estuarine water quality simulation system
concluding that the model is or is not verified. In a previous report such an analysis was made for the tidal flow in Jamaica Bay, using part of the data for model adjustment and the other part for actual verification. The subjective analysis was then extended by determining some statistical properties of the difference between observed and computed values.

This analysis has of necessity been extended again in this report. During the period when field measurements of water-quality parameters and rainfall were being made, the tide stage recorder in the bay malfunctioned. Without tide information, no meaningful simulation for an evaluation of both models could be made, and a method had to be developed to obtain the optimal estimate for the model input tide from observations made outside the bay. The results of this research effort, initiated by unfortunate circumstances, appeared to be fruitful and to have wide application, as they not only provide a way to make optimal estimates of missing records, but also are the basis of a novel way to adjust mathematical and physical hydraulic models to water-quality management and control as mentioned above.

Stochastic analysis has been applied to three different aspects of the study. First, it has been used for the extended analysis of observed and computed variables in the estuarine flow field in order to assess the behavior and resolving power of the numerical methods in the model and to obtain a quantitative evaluation of energy distributions. This research has considerably enhanced our understanding of the physical processes involved and has strengthened our high confidence in the numerical simulations.

Second, it was used to reconstruct the time history of water levels at locations where observed data are missing. The input tide for the model was thus reconstructed from relations between neighboring stations obtained at other times and from the extended analyses mentioned earlier.

Third, it has been used for determining the peripheral inputs of the urban drainage system overflow into the two-dimensional water-quality simulation model for the combined evaluation of both models; the results are presented in this report. However, the theory for these drainage basin models is discussed in a separate report. This report presents only the quality and quantity of rainfall-induced urban drainage overflow as used in the simulation.

In model evaluation the concept of state is used. Generally, we can refer to the state of certain quantifiable variables such as concentrations and water levels, although it is obvious that the true state is never known and that it is possible only to estimate the state of these variables.

Measurements of variables carried out in the bay do not represent the local state, but only an estimate of it, because of the characteristics of the instruments used and the quantity to be observed. Some variables can be measured accurately—for example, water levels. However, water-quality parameters such as coliform bacteria density can be determined only within a range of values. Thus a variable obtained by measurement contains a certain degree of error, which may even be influenced by the state before the observation time; the observation then presents a filtered estimate of the situation. Such is the case, for instance, with a tide gauge in which damping is introduced to filter out surface waves.

This report will present comparisons of the state estimates obtained from measurement and those from simulation. In the analysis extensive use was made of computer graphics; the more important results are presented in that form to enhance the intended message. However, we caution those readers without a thorough understanding of the physical and computational aspects of these methods not to derive independent conclusions from these graphs, since computational variations may at times cloud the physical phenomena. We have in general considerably
reduced the figures to present several on one page for easy comparison, although this impairs their readability.

Sections II and III contain theoretical aspects of our analyses; thus readers interested only in the description and the results of the simulation of the response of the bay to a rainstorm should begin with Section IV. Section II presents a general overview of the stochastic analysis method used in the investigations, and Section III is a discussion of the tide-induced hydrodynamic phenomena analyzed in the course of the study.
II. ANALYSIS IN THE FREQUENCY DOMAIN

This chapter will present the principles of the cross-spectral analysis methods used for the analysis of data obtained from observation and simulation and for the determination of system behavior. Complete theoretical developments will not be given, as they can be found in Goodman's thesis [1]; computational aspects are described by Liu [2].

LINEAR SYSTEMS

The method used in the analysis is based upon linear systems. Such a system has two groups of defined points of interest: One group can be called inputs, the other outputs. For simplicity, we will first consider a system with only one input and one output (Fig. 2). Ideal linear systems rarely occur, but in many instances linearity within certain ranges seems justified. It will be assumed that the characteristics of the system do not change during the time of our interest.

![Linear system diagram](image)

**Fig. 2—A linear system**

We are able to describe the dynamic characteristics of the system in two different ways, i.e., by an impulse response function \( h(\tau) \) or by a frequency response function \( H(f) \). Using the impulse response function description, the output \( y(t) \) of the system is given by the convolution integral

\[
y(t) = \int_{0}^{\infty} h(\tau) x(t - \tau) \, d\tau
\]  

Thus the value of output at \( t \) is obtained as a weighted summation over the entire history of the input \( x(t) \), with a weighting function \( h(\tau) \).

As its name indicates, the frequency response function \( H(f) \) describes the relation (in amplitude and phase) between a periodic fluctuation with a certain frequency \( f \) as input and the resulting fluctuation as output. Thus, if a linear system is excited by a periodic fluctuation \( x(t) = A \sin(2\pi ft) \), and a periodic fluctuation \( y(t) = B \sin(2\pi ft + \phi) \) is the result (Fig. 3), then the amplitude of the response function is

\[
|H(f)| = \frac{B}{A}
\]  

and the phase \( \arg[H(f)] = \phi \).
The frequency response function of the constant-parameter linear system is the Fourier transform of the impulse response function; thus

$$H(f) = \int_{0}^{\infty} h(\tau) e^{-i2\pi f \tau} d\tau$$  \hspace{1cm} (4)

The frequency response function is a complex variable; its absolute magnitude at a particular frequency is called the gain or amplification factor and the phase angle the phase factor.

The frequency response function $|H(f)|$ is often used in combination with the power spectral density function of the input and output.

The spectrum of a data set describes the general frequency composition in terms of the mean square value of each individual component. For example, if a record exists from a combination of two sinusoidal components

$$A_1 \sin (2\pi f_1 t) + A_2 \sin (2\pi f_2 t)$$  \hspace{1cm} (5)

then the spectrum would be two values on the graph. Such a spectrum is called a discrete spectrum or line spectrum (Fig. 4).

The line spectrum component is computed by integrating the squared components of the data set over its period and subsequently determining its mean value, as illustrated in Fig. 5.
If the data set (Eq. (5)) is the excitation of a system for which the frequency response function is known, then the resulting output spectrum can be obtained by multiplying the spectrum value at each frequency of the input with the square of the response function at that frequency (Fig. 6). Each component of the output is amplified and shifted in phase as shown in Fig. 7 before the mean square value is determined. It will be noted that in the spectral relations only the magnitudes are involved, not the phase relation between input and output.

In work with field data and real physical systems, inputs generally do not consist of a limited number of sinusoidal components. It is thus more appropriate to work with the concept of random or stochastic data for inputs and outputs. If we assume that the random data is stationary (by which is meant that certain statistical properties are invariable if sampled over sufficient time), then we can describe the data as being composed of an infinite number of sinusoidal components. The spectrum of the data then becomes continuous. Rather than plotting the total energy for a particular frequency, as is done for a line spectrum, one plots the spectral density, which can be thought of as the data set components’ mean square value within a narrow frequency interval between \( f \) and \( f + \Delta f \), divided by the frequency interval.

\[
P_{xx}(f) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \lim_{T \to \infty} \frac{1}{T} \int_0^T x^2(t, f, \Delta f) \, dt
\]  

(6)

In this equation, \( x(t, f, \Delta f) \) represents that part of the data set after all components not present in the frequency interval between \( f \) and \( f + \Delta f \) are removed. Thus the integral under the curve of the spectral density function represents the mean square value \( (\sigma_x)^2 \) of \( x(t) \)

\[
\int_0^\infty p_{xx}(f) \, df = (\sigma_x)^2
\]  

(7)
Fig. 6—Line spectrum of input and output and squared response function

\[ P_o(f_2) = |H(f_2)|^2 P_I \]

Fig. 7—Computation of line spectral component from input and output histories
As with the line spectra, the response function can be used to obtain the output spectrum from the input spectrum (Fig. 8).

\[ P_{yy}(f) = |H(f)|^2 P_{xx}(f) \]  \hspace{1cm} (8)

The mean square value of the output can be determined from the response operation and the input spectrum, as shown in Fig. 8.

In this figure the response function is zero for very low frequencies; as a result, the output at these frequencies is also zero, even though the input in these low frequencies actually exists. In this case, the system behaves as a so-called "high-pass filter."

It will be noted that in this relation between the three quantities only the magnitude of the frequency response function is involved, and no phase relations.

To determine phase relations between input and output, the concept of the cross-spectral density function is necessary.

---

**Fig. 8**—Input spectrum, response function, and resulting output spectrum
CROSS-SPECTRUM

The cross-spectrum also determines the relations between inputs and outputs, but includes the phase in this relationship. For a combination of sinusoidal components, a discrete cross-spectrum value is determined for each frequency. The cross-spectrum is a line spectrum, and is complex; thus both the real and the complex parts must be expressed. The real part of the spectrum, also called co-spectrum, is obtained from the average product of input and output in the same way that the input spectrum is obtained from the input, as was shown in Fig. 5. This computation for a single component is shown on the left side of Fig. 9. The quadrature line spectrum is the imaginary part of the cross-spectrum, and its value for a single component is similarly found from the input component and the so-called out-of-phase component of the output, which has a 90-deg phase shift in time compared to the in-phase component and the same amplitude. The representation of the computation of a single quadrature line spectrum value is also shown in Fig. 9.

![Diagram](image)

**Fig. 9**—Computation of co- and quadrature line spectral value from a single input and output component

In the case represented, the co-spectrum value is larger than the quadrature spectrum value. It will be noted that if the output is exactly in phase with the input, the quadrature spectrum value is zero.

In any case, the cross-spectrum value is expressed as a complex value.
\[ P_{10}(f) = C_{10}(f) = iQ_{10}(f) \] (9)

If polar notation is used, then we have the expression

\[ |P_{10}| = \sqrt{C_{10}^2(f) + Q_{10}^2(f)} \] (10)

and

\[ \theta_{10}(f) = \tan^{-1}\left(\frac{Q_{10}(f)}{C_{10}(f)}\right) \] (11)

It is customary to express the cross-spectrum as co-spectra and quadrature spectra rather than in polar coordinates, in contrast to the response function, which is always expressed in magnitude and phase.

In the discussion of spectral relations in the previous section, goniometric functions have been used, since these are somewhat more easily visualized. The mathematical manipulations needed to determine cross-spectral relations are more easily performed with complex functions. If the complex response function \( H(f) \) is expressed in polar notation, such as

\[ H(f) = |H(f)| e^{i\phi(f)} \] (12)

and the input component of the system is expressed as \( Ae^{i2\pi f_1 t} \), then the complex cross-spectrum value of frequency \( f_1 \) is computed as

\[
P_{10}(f_1) = \frac{1}{T} \int_{0}^{T} Ae^{i2\pi f_1 t} \cdot |H(f_1)| \cdot e^{i\phi(f_1)} Ae^{-i2\pi f_1 t} dt
\]

\[ = H(f_1) \cdot \frac{1}{T} \int_{0}^{T} [Ae^{i2\pi f_1 t}]^2 dt \] (13)

Thus

\[ P_{10}(f_1) = H(f_1) P_{1}(f_1) \] (14)

The input spectrum is real. Thus the phase angles of the response function and the cross-spectrum are the same; consequently,

\[ \phi = \theta \] (15)

where \( \phi = \) phase angle of the response function,

\( \theta = \) phase angle of the cross-spectral function.

Consequently, the response function value can be determined directly from the phase angle of the cross-spectrum. The absolute value of the cross-spectrum value is determined directly from Eq. (14):

\[ |H(f)| = \frac{|P_{10}(f)|}{P_{1}(f)} \] (16)
If we are working with real physical systems with random data, then we must again use the concept of spectral density.

The co-spectral density function can then be thought of as the mean of the product of input and output within a narrow frequency band $\Delta f$ and divided by the frequency interval

$$C_{xy}(f) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t, f, \Delta f) y(t, f, \Delta f) \, dt$$  \hspace{1cm} (17)$$

The quadrature spectral density function is then thought to be the mean of the product of input $(x)$ and $(y^*); the latter is shifted in phase 90 deg from $y$, within a narrow frequency band $\Delta f$ and divided by the frequency interval

$$Q_{xy}(f) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t, f, \Delta f) y^*(t, f, \Delta f) \, dt$$ \hspace{1cm} (18)$$

The absolute value of the cross-spectral density function is

$$|P_{xy}(f)| = \sqrt{C_{xy}^2(f) + Q_{xy}^2(f)}$$ \hspace{1cm} (19)$$

and the phase of the cross-spectral density function is

$$\theta_{xy}(f) = \tan^{-1}\left(\frac{Q_{xy}(f)}{C_{xy}(f)}\right)$$ \hspace{1cm} (20)$$

The cross-spectral density function is related to the input spectrum and the response function as follows:

$$|P_{xy}(f)| = |H(f)| \cdot P_{xx}(f)$$ \hspace{1cm} (21)$$

$$\theta_{xy}(f) = \theta(f)$$ \hspace{1cm} (22)$$

Thus, if the cross-spectral density function and the input spectrum are known, the amplitude as well as the phase of the response function can be determined. The relation between the cross-spectral density function and the input spectrum is illustrated in Fig. 10.

**UNCORRELATED COMPONENTS**

In studying real systems, the inputs and outputs often cannot be measured accurately; consequently, only estimates of the spectral density functions can be obtained. However, cross-spectral analysis also provides a means for estimating the properties of the measurement errors.

Assume a linear system in which the input is determined exactly but in which the output has a certain random error caused (for example) by the instrument used
Fig. 10—Relation between the input spectral density function and the cross-spectral density function

for recording the output. The measured output's deviation from the actual output, which is often called noise, also has a spectral density function, and the measured spectral densities are those from the system's output and the noise of the instrumentation combined. Consequently, the measured output spectrum overestimates the actual output, and application of Eq. (8) overestimates the response function amplitude.

The numerical method for estimating the response function from the cross-spectrum by use of Eq. (21), as described by Liu [2], is an optimum estimate of the response function.

The ratio of the estimated response function amplitude obtained from the cross-spectral density function (Eq. (21)) and the response function amplitude from the input and output spectra relation (Eq. (8)) is called the coherency. For the squared coherency function we find from the estimated spectra \( \hat{P}_{xx}, \hat{P}_{yy}, \) and the estimated cross-spectral density, \( \hat{P}_{xy}, \)

\[
\phi_{xy}^2(f) = \frac{[|\hat{P}_{xy}(f)|/|\hat{P}_{xx}(f)|]^2}{|\hat{P}_{yy}(f)/\hat{P}_{xx}(f)|} = \frac{|\hat{P}_{xy}(f)|^2}{P_{yy}(f) P_{xx}(f)}
\]  

(23)
The coherency for an ideal linear system with no noise in measured input and output is unity. The coherency is an indicator of the system behavior. If the system is somewhat nonlinear, those parts of the output are not correlated with the input and are reflected in the coherency function in a manner similar to that described above.

In the determination of the behavior of a system it is often useful to see how the noise is distributed over the frequency range. The estimated spectral density function of the noise is expressed by

\[
\hat{P}_{xy} (f) = \left[ 1 - \hat{\alpha}_{xy}^2 (f) \right] \hat{P}_{yy} (f)
\]

(24)

In Fig. 11 an example is presented of the spectral relations of a system with uncorrelated components in its output.

The computation of the different estimates of spectra, cross-spectra, and other significant functions, such as the response function from data sets, requires rather extensive calculations, not presented here. This section has presented an overview of the physical concepts of spectra and their related functions used in the analyses presented in the next section. The method used for the actual computation of estimates, described by Liu [2], uses uninterrupted data sets sampled at equally spaced time intervals. This method can be extended for data sets with missing data, as described in Appendix A.

Fig. 11—Spectral relation of a system with uncorrelated components in its output
III. CROSS-SPECTRAL ANALYSIS OF TIDES AND TIDAL FLOW

In this section cross-spectral analysis will be applied to different sets of data to investigate phenomena both measured in the field and computed with the model.

ENERGY DISPERSION

In the Jamaica Bay modeling effort, much emphasis has been placed upon a highly detailed representation of the tidal flow. If the flow can be represented in much detail by using a fine grid, then the major part of the transport of the constituents is represented by advection. Only motions of the scale of the grid size are represented by dispersion (thus by parametric expressions). Consequently, the state estimates can be expected to be much more accurate than those obtained by models using parametric expression nearly exclusively.

If the fluid motions in the estuary are two-dimensional and have isotropic density, as we have assumed in the model, then we can expect that long waves penetrating the estuary from the ocean will not only change in amplitude but also transfer energy from one wave frequency to another. This frequency dispersion process is very complex, but we are able to characterize some of the main mechanisms by which it occurs.

If a wave, like a tidal wave, propagates through an estuary in one dimension with constant depth, the nonlinear advective term $u \delta u / \delta x$ in the equation of motion and the nonlinear transport term $\delta (Hu) / \delta x$ generate higher harmonics of the fundamental frequency. In particular, the second harmonic increases in amplitude with the distance of propagation. The tides generated by this process are called tides of second order. Higher order terms are also generated, but these are generally not predominant. Similarly, the bottom friction and a sloping bottom generate harmonics as described by Gallagher and Munk [3]. The bottom friction generates odd harmonics of the frequency.

Other mechanisms of energy transfer in the frequency domain can be found which are related to the two-dimensionality. The varying depth and complicated boundaries generate dispersion of energy in frequency by a combination of effects described above and by the generation of local circulations through advection.

The latter is of particular interest in our modeling effort, since such local circulations may be of considerable importance for the dispersion of constituents. These circulations are indirectly tide-induced by a mechanism illustrated in Fig. 12. In estuaries with irregular boundaries this may be one of the major mechanisms by which pollutants are dispersed. Such local circulations can be considered two-dimensional isotropic turbulence, which would then exist simultaneously with the three-dimensional isotropic turbulence on a much smaller scale. These local two-dimensional circulations become established by the flood and ebb currents in the main channels of the estuary. During slack tide, these circulations continue by inertia and radiate their energy in part into the main body of water.

Also, the spatial variability of the velocities existing during ebb and flood in estuaries with an irregular bathymetry will drastically change near slack and induce local circulations and water level oscillations. During this period the currents
can be characterized as being weak and variable, but they are of great importance in the mixing process.

If several frequencies are present simultaneously, we can also expect, because of nonlinearity, generation of frequency sums and differences as the tide travels through the estuary. For example, if two waves are present running in $x$ direction with frequencies $\omega_1$ and $\omega_2$ and amplitudes $u_1^*$ and $u_2^*$, the velocity $u(t,x)$ can be expressed by

$$u(t,x) = u_1^* \sin (\omega_1 t + \sigma_1 x) + u_2^* \sin (\omega_2 t + \sigma_2 x)$$

and the velocity gradient by

$$\frac{\partial u}{\partial x} = u_1^* \sigma_1 \cos (\omega_1 t + \sigma_1 x) + u_2^* \sigma_2 \cos (\omega_2 t + \sigma_2 x)$$

Consequently, the advection term $u \frac{\partial u}{\partial x}$ generates, besides waves with frequencies $2\omega_1$ and $2\omega_2$, waves with the difference in frequency $\omega_1 - \omega_2$ and the sum of the frequencies $\omega_1 + \omega_2$, as can easily be derived from Eqs. (25) and (26) by multiplying the right-hand sides and expanding the trigonometric functions. Thus energy can be transferred from one frequency to higher and lower frequencies by interaction of waves with different frequencies. The question immediately raised now is, How does the model represent the frequency dispersion, and how are we able to make comparisons between the prototype and our model?

As expressed in a previous volume of this series, the observations of the flow in Jamaica Bay have been primitive, and consequently cannot be used for any detailed analysis other than for a quick comparison between observed and computed velocities. No details of the internal flow structure are available, since integrated values or mean values are used over a period of one minute, which will then be approximately half an hour apart. The water level observations, however, are quite different. These were obtained by instruments which record the water level at intervals of 0.01 ft and at 6-min time intervals. The locations of the recording gauges for the survey described in Ref. 4 are indicated in Fig. 13. From these time series it is possible to compute spectral density distributions of the tidal heights as a function of frequency. Figure 14 represents part of the results of the spectral computations for the observed and computed water levels at Rockaway and Kennedy. The record length is 75 hr, from an observation made in October 1970, described in Vol. IV of
Fig. 14—Time series of observed and computed water levels at Rockaway and Kennedy, their spectra, impulse response function, coherency, and noise spectra.
this series. Only every fifth value of the record was used in the analysis. In the Kennedy records a difference in the mean level can be noted, since no correction was made for the deviation of the datum for the Kennedy gauge. It will be noted that observed and computed spectra have very small differences. The observed spectral density estimate for Rockaway is larger than that computed.

For our simulation we used as input for the model the observations from the Rockaway gauge, which were subsequently smoothed and corrected for the effect of the resonance mode in the model. The smoothing operation is probably the reason that the spectral density estimate of the model at Rockaway is lower than the observed at that location. Nevertheless, the agreement between observed and computed is good, and the difference in the spectral density estimates is only a few percent.

The spectral computations discussed here were based upon the observed tide records. In contrast to the representations in Ref. 4, corrections for the advance or lag of the timing mechanism of the stage recorders have been made, as described in Appendix B. These corrections reduced the standard deviation between observed and computed tidal heights to about 0.05 ft, and now no substantial difference exists between the standard deviations for the adjustment period and verification period. Table 1 presents the standard deviations between observed and computed water levels at the condition for which we considered that no further bottom friction adjustments were required.

<table>
<thead>
<tr>
<th>Tide Gauge Station</th>
<th>Oct. 27 and 28 Adjustment Simulation</th>
<th>Oct. 29 and 30 Verification* Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rockaway</td>
<td>0.032</td>
<td>0.033</td>
</tr>
<tr>
<td>Rosie's Boatyard</td>
<td>0.043</td>
<td>0.050</td>
</tr>
<tr>
<td>Canarsie Pier</td>
<td>0.034</td>
<td>0.038</td>
</tr>
<tr>
<td>Kennedy</td>
<td>0.050</td>
<td>0.065</td>
</tr>
<tr>
<td>Head of Bay</td>
<td>0.064</td>
<td>0.081</td>
</tr>
<tr>
<td>Average</td>
<td>0.045</td>
<td>0.053</td>
</tr>
</tbody>
</table>

*Comparison is made after correction for effect of slow Rockaway clock and fast Kennedy clock (see Appendix B).

It can be noted in Fig. 14 that the coherency between the records is close to unity over major portions of the frequency range. The $M_2$ tide is well correlated.

From the spectra of the uncorrelated components it can be seen that the density of the noise at the $M_2$ tide frequency is less than 0.1 percent of the density of the spectra at the stations. Since nearly all the energy is in the low frequencies, the amplitude of the uncorrelated components should consequently be only a few percent of the tidal amplitude and is comparable with the standard deviation between
observed and computed data. It will be noted also that the computed and observed noise spectra are comparable; thus, the model generates the uncorrelated components by the nonlinear expressions in the equations and by the complicated bathymetry rather well.

Figure 15 presents the frequency response function between Rockaway and Kennedy. It will be noted that the tidal amplification between the station for the $M_2$ tide is a few percent too high in the computation. This is also the case in the higher frequency range. The time lags do agree extremely well and are generally within the accuracy range of the clocks of the gauges (1 min).

In the analysis of the computation scheme in Ref. 5, the propagation factor was introduced to assess the properties of a computation scheme in relation to the real solution of the linearized equations. The propagation factor, expressed in the dimensionless parameter $(\sigma L)$, was defined as the complex ratio of the computed wave (in amplitude and phase) to the physical wave after a time interval in which the physical wave propagates over its wavelength. Figures 16 and 17, taken from Ref. 5, present phase angle and modules of the propagation factor for typical linearized damping.

The discreteness of the spatial representation is presented in parts per wavelength. Thus, a wave with a wavelength of 50,000 ft computed with a grid size of 1,000 ft has 50 parts per wavelength. The discreteness in time is presented as a function of the propagation speed of the wave in $\tau \sqrt{gh}$. The amplitude propagation factor of the computational scheme is always larger than one, while the speed lags.*

Thus, in our model, when the waves travel over a fixed distance, the computed wave should be slower than the observed wave. This is indeed observed for the response function at the frequency of the $M_2$ tide, as shown in Fig. 15.

**USE OF RESPONSE FUNCTION FOR MODEL ADJUSTMENT**

In the adjustment of the model as described in Ref. 4 the best fit of observed and computed data was obtained. This resulted in particular emphasis on the phase relation between observed and computed data, rather than on amplitude. Because the shorter waves are more affected, the computed response function has a higher value than the observed.

As described in Ref. 4, the Manning's coefficient used in the simulation was $n = 0.030$ over the whole computational field, except in limited areas where it was taken to be $n = 0.034$. To investigate the effect of bottom friction on phase and amplitude of the tide, an experiment was made with a minor change in the Manning's coefficient. In this experiment the coefficient was taken to be $n = 0.031$; in the areas where $n = 0.034$ was used this value was maintained. The frequency response function of this simulation is presented in Fig. 18. Compared with the frequency response function in Fig. 15, the amplification of the $M_2$ tide changed only insignificantly, and the phase lag between Rockaway and Kennedy increased about 3 min. The larger lag at Kennedy increased the standard deviation between observed and computed from 0.050 ft to 0.065 ft for the model adjustment period of October 27 and 28 for this particular experiment with $n = 0.031$.

Interestingly, the theoretical analysis of the computational behavior of the model in Ref. 5 (using the propagation factor) concluded that because of the approximations of the finite-difference scheme, considerable adjustment difficulties are encountered.

Satisfying both amplification and phase for every frequency by adjusting bottom

* This is based upon an analysis of a one-dimensional system. The wave amplification in a two-dimensional model is not necessarily the same as in a one-dimensional model.
Fig. 15—Frequency response function between Rockaway and Kennedy from observation and computation

Fig. 16—Phase angle of the propagation factor of the computational method
Fig. 17—Modulus of the propagation factor of the computational method.

Fig. 18—Frequency response function between Rockaway and Kennedy from observation and computation with a Manning's value of $n = 0.031$. 
friction is impossible. It can be achieved only for one frequency and then only by also making adjustments in depth. These considerations are theoretical; in practice a satisfactory agreement between computed and observed data can be reached, but the spatial discreteness intervals chosen should be very small.

For the adjustment of a model we are concerned not only with tidal amplification, but also with circulation. In the case of Jamaica Bay, a net circulation is generated by the propagation of the tidal wave through the channel system. The channels are quite pronounced. Because many are dredged, their slopes are steep, which is evident when the 5-ft and 15-ft depth contours are plotted in the model (Fig. 19).

For proper simulation of circulation, the crosswise gradients in prototype and model should be the same, since this is the driving force generating the weak net currents in Broad Beach Channel and Grassy Bay (Fig. 13). In the analysis of the crosswise gradient, the tide stations at Canarsie and Rosie's Boatyard are used, which are about the same distance from the channel divergence point near Rockaway Inlet.

Figure 20 presents the results of the spectral computation. The coherency between the two observed and computed time series is nearly unity over the entire frequency range. The spectral density of the uncorrelated components is four orders of magnitude smaller than that of the tide records at the frequency of the $M_2$ tide. Thus both branches of the system do behave very similarly as to generation of nonlinearities.

The response function (Fig. 21) shows excellent agreement for the $M_2$ tide. The crosswise amplification is nearly unity for the observation and simulation. The difference in phase lag at the $M_2$ frequency between computed and observed is about 0.5 min.

In the discussion of Ref. 4, reference was made to a physical model of Jamaica Bay built primarily to assist in studies of hurricane surge protection of the area. From experiments with that model, it appeared that the circulation was clockwise. Results with the present model, however, indicate a counterclockwise circulation. In the analysis of water-quality distributions described in the next section it is also shown that the circulation is counterclockwise. We have concluded that the physical model should be adjusted further for purposes of water-quality research (in which the simulation of circulation is an important aspect), since the physical model was adjusted on much less accurate data than was available for this effort. Consequently, it could be expected that the response function of the model observations at Canarsie and Rosie's Boatyard, made before such adjustments for water-quality research were made, would be significantly different. This is particularly true for the phase as shown in Fig. 22, where the phase relation of the numerical simulation is also presented.

**DETERMINING SYSTEM BEHAVIOR BY SPECTRAL METHODS**

Stochastic analysis determines from two time series their interrelation or lack of interrelation. This section investigates the relationship between tide elevation and tidal currents at two stations in the bay. During the field observations only short-duration velocity records were obtained, which were only about 14 hr in length.

*This adjustment has recently been made. The response functions of the hydraulic model are now in excellent agreement with prototype data. A separate report on this analysis is now being prepared.*
Fig. 10. The 5 and 15 ft. depth contours in the Jamaica Bay model.

JAMAICA BAY

500' F.O.
15.0.
Fig. 20—Time series of observed and computed water levels at Canarsie and Rosie's Boatyard, their spectra, impulse response function, coherency, and noise spectra
Fig. 21—Frequency response function between Canarsie and Rosie's Boatyard

Fig. 22—Time lag spectra of the numerical simulation model and the physical model compared to timelag spectra from observation
each and taken at half-hour intervals, which is too short for an analysis with reasonable confidence. The analysis was also made using a numerical simulation of the tide on May 31, 1972 through June 2, 1972, details of which will be given in Section IV. For the purposes of analysis, water levels and currents at both Canarsie and Rosie's Boatyard were abstracted from the model results every 6 min. Only 50 hr of real time could be used effectively for analysis. The results of the spectra computations are shown in Fig. 23.

In the velocity spectra, besides a peak at the \( M_2 \) tide frequency, a second peak is present at a frequency of \( 0.24 \text{ hr}^{-1} \). These peaks in the spectra are odd harmonics of the basic frequency and are attributed to bottom friction.

The coherency spectra show coherency in three ranges, namely, near the \( M_2 \) tide frequency, around \( 0.3 \text{ hr}^{-1} \), and \( 0.5 \text{ hr}^{-1} \).

The response function is presented in Fig. 24. At the \( M_2 \) tide frequency the currents are about 90 deg out of phase with the tide levels. With increasing frequency, the lag decreases, and around \( 0.4 \text{ hr}^{-1} \) the lag becomes an advance. Around this frequency a nodal point should exist near Canarsie and Rosie's Boatyard, which agrees with the low coherency found. At a nodal point of an oscillating basin, no water level fluctuation at the oscillating frequency is present. At that point currents do exist, however; thus no relation between currents and water levels exists.

Of interest also are the oscillations with periods of about 2 hr. At this frequency, the coherency is unity, a peak exists in the amplitude of the response operator, and the phase advance is about a quarter of a circle (90 deg).

In the plot of the amplitude of the response function a low peak is present at \( 0.3 \text{ hr}^{-1} \). This coincides with the high coherency at that frequency and can be related to the basic oscillation of the bay with the node near the inlet.

In comparison with the response functions between two tide stations, the response function between water level and velocity is markedly different. The coherency is generally much lower, which indicates that the system is quite far from being linear for the currents.

**ENERGY DISTRIBUTIONS IN HIGHER FREQUENCIES**

Of particular interest are the distributions of the spectral density for the higher frequencies from the simulation, which are presented in Fig. 25 for Canarsie and Rosie's Boatyard for computed and observed data. It will be noted that the spectral density for frequencies with periods shorter than one hour is nearly a straight line when plotted on our log-log scale, and that the energy decreases according to a \(-5/3\) power of the frequency. One could conclude that the potential energy spectrum agrees with the Kolmogorov spectrum law [6] for the inertial subrange in which energy cascades from lower to higher wave numbers according to a \(-5/3\) power. Several investigators have observed this frequency decay for the kinetic energy spectra [7,8].

However, some doubt exists as to this conclusion. The spectral analysis method used has a so-called Tukey window [2] because of the desire to establish the spectral estimates with high accuracy. This spectral window has so-called leakage, by which energy of a high-intensity frequency band is transferred into other frequency bands. Thus the estimated spectral representation is contaminated by this leakage. Consequently, an analysis of this type can be made more effectively by first removing the \( M_2 \) tide from the record, then making the analysis.

Since leakage is a secondary effect, the computed spectra would be affected in about the same magnitude as the observed spectra. Thus it is apparent that the
Fig. 23—Time series of computed water levels, velocities, spectra, and cross-spectra between water levels and velocities at Canarsie and Rosie’s Boatyard.
Fig. 24—Frequency response functions between water levels and velocities at Canarsie and Rosie's Boatyard

Fig. 25—Computed (•) and observed (o) spectral estimates derived from observed and computed water levels at Canarsie and Rosie's Boatyard
potential energy disperses in the model as in nature, because the computed spectral density values agree well with observed values.

Earlier in this section some of the mechanisms of energy dispersion have been described. Generally the energy transfer can be attributed to turbulence.

Three-dimensional turbulence cannot be directly simulated by the model, as the vertical velocities do not appear in the formulation of the model; it is simulated by using parametric expressions for bottom stress. Simulation of two-dimensional turbulence is possible, however, and the existence of smaller and larger two-dimensional vortices can be observed in the model. Naturally, only vortices equal to or larger than the grid dimension can be simulated. Smaller vortices cannot be generated, nor can waves be generated with wavelengths smaller than twice the grid size.

In prototype the energy transfer is continuous until final dissipation by viscosity at very high wave numbers, while a limit exists in the model. If proper precautions are not taken in the design of the computational scheme, the cascading energy will accumulate in very short waves with wavelengths of twice the grid size and in small vortices, giving rise to nonlinear instability of a computation. This instability is manifested as rapidly growing alternating high and low values of the water levels, and as "noodling" of the velocity field (Fig. 26). In the latter case the weaker velocity component (thus the one which is nearly perpendicular to the current) will alternate in direction from gridpoint to gridpoint.

To prevent this effect, one can design a scheme which dissipates energy in the higher frequencies, or one can design a scheme which conserves quadratic properties such as potential energy \( 1/2gh^2 \), kinetic energy \( 1/2v^2 \), and squared vorticity \( (\partial u/\partial y - \partial v/\partial x)^2 \).

It seems inappropriate to rely only on conservation of quadratic properties mentioned above. If this were applied, it would mean that in the absence of bottom friction (related to the local velocity squared), all energy would be returned within the spatial and time representation of the computation scheme, which is naturally not the case. Thus energy should be dissipated out of the higher wave frequencies and higher wave number region of the model. In the Jamaica Bay model simulations

![Fig. 26—Unstable velocity field (noodling)](image-url)
this is accomplished by periodic time smoothing and by the use of a quadratic bottom
friction term. By the first procedure, every few hours (real time) the computation
was restarted using the averaged values of velocity components and water levels
over one time step, with proper allowance made for the time shift of half a time step
at that time. This takes short waves in the time domain out of the simulations.
Alternatively, a viscosity term has been used with the same results. It can be shown
that the computational method used in the simulations also conserves certain quad-
tratic properties.

Since Jamaica Bay has a channel structure, noodling was possibly inhibited by
boundaries. In wide and open estuaries this nonlinear instability of the flow field is
much more possible; to make the model more generally applicable, a study of the
stochastic properties of the prototype flow field at spatial intervals of the simulation
of such an estuary will be required before a conclusion can be reached about impro-
vising the simulation method by other advection expressions. In any case, in such
models of wide estuaries both bottom stress and viscosity must be adjusted.

RECONSTRUCTION OF A MISSING RECORD

During the period of observation of the spatial and temporal distributions of
coliorm bacteria described in the next section, the tide gauge in the bay malfunc-
tioned without being noticed during the field sampling. Without tide information,
no meaningful simulation could be made, since the tide record was to be used for
deriving the input tide at the open boundary of the model. Rather than request-
ing another survey, we decided to reconstruct the missing record, particularly since all
other data appeared to be good and the type of rainstorm was adequate for analysis.

For the reconstruction we had available two data sets (Fig. 27), i.e.,

1. Observed tide data at Canarsie Pier during a 58-hr period from May 29 to
May 31, 1972. As can be seen in Fig. 27, the gauge started malfunctioning
on May 31 at 9:00 a.m. This data was recorded digitally at 15-min intervals.
2. Observed tide data at East Rockaway (Fig. 27) from May 29 to June 6, 1972.
The observed data was continuous and obtained from a Bristol gauge in-
stalled and operated by the U.S. Geological Survey. In Fig. 27 a 6-day record
is shown for the period from May 29 to June 4, 1972.

It will be noted that the tide gauge at Canarsie was operative for approximately
58 hr. Consequently, it seems possible to obtain the missing record of Canarsie by
making a Fourier series decomposition of the record and then, by use of the compo-
nents, extending the series in time. The result thus obtained is not an optimum
estimate, since only one record of limited time is used and the contributions of the
random component introduced by major meteorological disturbances are unknown.

By making a cross-spectral analysis of the two available records, we are able to
determine the optimum relation between these observations. The results of the
spectral analysis are presented in Figs. 28 and 29. For this analysis a record of 50
hr was used, as this is a multiple of the $M_2$ tide period. It will be noted that the
spectral estimates of the uncorrelated part are about one order of magnitude higher
than those obtained in previous cross-spectra computations. This can be explained
by the shorter observation period, the errors introduced by digitizing an analog
record, and by the larger distance between the two stations.

For the reconstruction, the whole record of the East Rockaway station was then
decomposed for the same frequencies as the frequency response function estimate.
Subsequently, using the frequency response function, we were able to compute all amplitudes and phases of the tide at Canarsie.

However, our interest is not the reconstructed tide record at Canarsie, but the tide record at Rockaway in the inlet of Jamaica Bay, since the latter data is essential for driving the model for the simulation of the tide and coliform distributions during a rainstorm from May 31 through June 2, 1972.

The tide levels at Canarsie and Rockaway were recorded in the period from October 27 to October 31, 1970, as described in Volume IV of this study and presented in Fig. 30.

The cross-spectral analysis of these records (Figs. 31 and 32) shows a high coherency; information can be transferred from one station to another with much confidence. For the reconstruction of the tide at Rockaway, the amplitudes and phases were cascaded from East Rockaway to Canarsie and from Canarsie to Rockaway. For the water-quality simulation of the period from May 31 through June 2, 1972, the tide components with periods shorter than one hour were purposely omitted. We assumed that these components would not be essential in the simulation because of their very low density. The intent of this omission was to investigate the generation of these components by the nonlinear behavior of the model.
Fig. 28—Observed water levels at Canarsie and East Rockaway, their spectra and response functions
Fig. 29—Coherency spectrum, noise spectrum, and impulse response function between Canarsie and East Rockaway

Fig. 30—Observed water levels at Rockaway and Canarsie for the period of October 27 to October 30, 1970
Fig. 31—Time series of observed and computed water levels at Rockaway and Canarsie, their spectra, impulse response function, coherency, and noise spectra
Fig. 32—Frequency response functions between Rockaway and Canarsie from observation and computation

In the graphs of the spectral analysis of the observed data between Rockaway and Canarsie, the results of the spectral analysis of the simulation described in Ref. 4 are also shown. Note that the uncorrelated components (noise) have about the same intensity. The computed response function amplitude is slightly larger than the observed amplitude, and the time lag is slightly higher in the model than in the prototype, as would be expected.

The results of the optimal state estimate of the tide level at Rockaway for the period of May 31 through June 3, 1972 are presented in Fig. 33, together with the observations at East Rockaway. Late in the evening of May 31, 1972, the rainstorm started, which induced a small disturbance in the record of East Rockaway; this disturbance is also generated in the tide to be used for simulation. We will see that this perturbation will be reinforced in most simulated tide records in the bay by the large amount of water added to the system by the rainstorm.

The comparison of the two tides in Fig. 33 shows another interesting feature. The amplification shows up mainly at high water (HW), which agrees with the theory of long wave amplification in shallow water.
Fig. 33—Observed water level at East Rockaway, together with reconstructed water level at the model boundary.
IV. STATE ESTIMATION OF COLIFORM DISTRIBUTION
BY OBSERVATION AND SIMULATION

This section presents a comparison of the estimated coliform distributions observed in Jamaica Bay following a rainstorm and the estimates obtained by simulating the effect of this rainstorm. The spatial and temporal distributions of coliform were selected primarily for the following reasons:

1. Coliform is the indicator organism used by public health organizations to determine the suitability of water for various types of human use.
2. If model adjustments are made using coliform distribution rather than some other constituent, then the model has a higher reliability for coliform, permitting more accurate future coliform management analyses.
3. Because of the relatively high dieoff rate of coliform, the background (system memory) at the beginning of a storm is insignificant, in contrast with other water-quality parameters, where the distributions in the bay may be influenced by a previous storm.
4. The urban drainage models provide input to the estuarine water-quality model; the coliform concentration has, among all the constituents, the highest predictability provided by these urban drainage models, second only to the overflow quantity predictions.

OBSERVED INPUTS AND OTHER PARAMETERS

During the water-quality simulation period (May 31, 1972 through June 2, 1972) observed data were made available by the City of New York and other agencies. The locations of the observations are shown in Fig. 1. Observed parameters used in the simulation and the analysis thereof include the following:

1. Hourly wind measurements at Kennedy International Airport.
2. Rainfall data measured at the 26th Ward WPCF and Kennedy International Airport. (Gauges at St. Albans and the Paerdegat pumping plant were inoperative.)
3. Water level data measured at Canarsie Pier and the U.S. Geological Survey gauge located at East Rockaway, which is an inlet southeast of Jamaica Bay.
4. The quantity of effluent from all four WPCFs and the quantity and coliform concentration of the bypassed flow from the 26th Ward WPCF.
5. Coliform concentrations at five stations in the northwestern part of Jamaica Bay. These stations are identified as stations 7, 8, 9, 10, and 14.

The wind, rainfall, and tide data, together with the observations of the WPCFs, were the inputs for the simulations. The measured coliform concentrations in the bay were used for comparison with the simulation. Graphic representations of the data used for the input are presented in Fig. 34.

During the field observations described above, water temperature and salinity were also measured at the five stations in the bay. This information was not used in the simulation or the analysis thereof.

Because not all data could be used directly in the water-quality simulation,
Fig. 34—Data used for input of the models
transformations had to be made at great effort. The generation of the input tide at Rockaway was discussed in Section III of this report, and the determination of the quantity and quality inputs is described below.

PERIPHERAL INPUTS

For the simulation of water quality in the bay the quantity and quality of the input at each of the discharge points must be determined. A few of these discharges, i.e., the inputs of the WPCFs and their bypasses, were measured and are directly available for use in the simulation. The other discharges are mainly storm-induced. For each of the drainage basins shown in Fig. 1, models were developed that use a response function to determine overflow quantity from rainfall. The response functions were based upon extensive measurements of rainfall and overflow for five of the basins. For the other basins, results of adjacent basins were used with adjustments. In determining the response function, extensive use was made of cross-spectral analysis. A description of the principles of these models is presented by Liu [2].

For each of the drainage basins the time history of rainfall must be determined before the overflow quantity and quality can be computed. In the design of the field measurement program for this experiment, four rainfall measuring stations were incorporated in order to form a network from which an estimate could be made in each basin as to the time histories of a moving storm.

Three stations are sufficient to obtain an estimate of a moving storm for a two-dimensional transfer using a two-dimensional spectral analysis of records [2]. Four stations were used to meet the contingency of failure of one gauge. Unfortunately, two gauges became inoperative, thus making an estimate using this technique impossible. Thus the overflow quantity and quality for this simulation were determined by using the rainfall data of the closest rain gauge station as input for the models of the drainage basins.

The variability in intensity from one station to another is considerable (Fig. 34). The rain starts at about 2000 hr at the Kennedy station and at about 2200 hr at the 26th Ward station, which is about 5 mi west of the former. The total rainfall at the 26th Ward was 0.28 in.; the rainfall at Kennedy was nearly 50 percent higher, namely, 0.41 in.

Figure 35 presents model results of the response of three urban drainage basins to the rainstorm. For each system three graphs are presented. For example, for the Thurston combined sewer system, which discharges through an overflow, the top graph shows the hyetograph and the resulting system overflow graph; they are both expressed in terms of inches per hour per unit area for easy comparison. Because the tilting bucket type of rain gauge (0.01-in. bucket) has relatively low resolving power for the prolonged and less intense storm, the computed rainfall intensities contain only steps.

It will be noted that the total rainfall quantity is much larger than the overflow, as there is considerable loss from infiltration. Also, part of this precipitation will be handled by the WPCF as intercepted flow.

The middle graph shows the resulting system overflow hydrograph in terms of cubic feet per second for the input to the two-dimensional water-quality simulation. The lower graph shows the computed time history of coliform concentration in the overflow. These transient pollutant emission estimates were computed from a formula that incorporates the basic mechanisms of surface washing, conduit dilution,
Fig. 35—Response of the Thurston Basin overflow, the Hendrix system overflow, and the Paerdegat system overflow to the rainstorm of May 31, 1972
and bottom scouring in the sewers [2]. The parameters in this formula must be
deduced from field data. Because of the characteristics of this particular storm—i.e.,
a low intensity and prolonged duration—the resulting coliform density in the emis-
Sion is in essence due to dilution of storm runoff with the sanitary flow, and the effect
of surface washing and bottom scouring is small.

The Hendrix Creek system is similar, and is presented in the same manner. The
Paerdegat system has a different behavior. This system is controlled downstream,
and starts discharging only after the water level within the system exceeds the level
in the estuary. The histories of water levels on each side of the tide gate are present-
ed in the middle graph. This system is quite large (see Fig. 1) and consequently slow
to respond. Inside the system, water levels start to rise after about three hours, and
only after six hours from the start of the rainstorm is overflow initiated.

The discharge into the bay presented in the bottom graph is of short duration,
and shows a peak (pulse) at the end. This peak is generated by the computational
procedure, and should be disregarded. Because of the large ponding effect of this
system, coliform concentrations become quite uniform during the discharge period.
Average values obtained from measurement of a large number of rainstorms were
used as input for the simulation.

Figure 36 presents the results of the Rockaway drainage basin and also the small
drainage basins of the residential areas on the island in the bay. The representa-
tion is similar to the one for the Thurston Basin system. The top graphs present the
hyetograph and the overflow graph in inches per hour, the middle graphs the system
overflow in cubic feet per second, and the bottom graphs the computed coliform
concentrations.

Figure 37 presents the results of the drainage basin response to the rainfall for
the two Spring Creek Basins and the Bergen Basin combined sewer overflow (CSO).
The discharges of the two Spring Creek systems are now intercepted by the new
Spring Creek AWPCF. The computations indicated that the facility was able to
retain the complete discharge, and observations confirmed this conclusion. Conse-
quently, no inputs were required for the Spring Creek Basins in the water-quality
simulation of the bay. The Bergen Basin discharge from this overflow was consider-
able, comprising the largest total load of the system.

For the separate storm water systems such as Thurston, Mill Basin, Kennedy
East and West, and Bergen Basin, the discharge graphs were also computed by the
urban drainage basin models. The qualities, however, were estimated from earlier
studies for the City of New York. These studies, based upon field data monitoring
during a summer season, showed that a flow-weighted mean value of 9000 MPN/ml
is an adequate estimate.

From field data we also derived inputs for the leachate of the Fountain Avenue
landfill (a coliform source concentration of 2400 MPN/ml) and the Rockaway landfill
(390 MPN/ml).

A total of sixteen inputs into the model (Fig. 1) were used:

1. 26th Ward WPCF effluent
2. Jamaica WPCF effluent
3. Rockaway WPCF effluent
4. Hendrix combined system overflow
5. Thurston combined system overflow
6. Bergen combined system overflow
7. Paerdegat combined system overflow
8. Fresh Creek combined system overflow
9. Rockaway combined system overflow
Fig. 36—Response of the Rockaway system overflow, Broad Channel West overflow, and Broad Channel East overflow to the rainstorm of May 31, 1972
Fig. 37—Response of the Spring Creek West system, the Spring Creek East system, and the Bergen combined system overflow to the rainstorm of May 31, 1972.
10. Bergen separate system (storm water) overflow and surface runoff from the west portion of the Kennedy International Airport, discharging at the end of Bergen Basin
11. Thurston separate system (storm water) overflow and surface runoff from the east portion of the Kennedy Airport, discharging at the end of Thurston Basin
12. Mill separate system (storm water) overflow
13. Runoff from the Fountain Avenue landfill
14. Runoff from the Rockaway landfill
15. Broad Channel East combined system overflow
16. Broad Channel West combined system overflow

As can be seen in Fig. 1, the drainage basins along the northern boundary of the bay do not extend to the shore. These areas between the drainage basins and the shore are generally uninhabited, with low social activity; no inputs were made for these areas. Also, inputs near the center of the bay, with its large bird population, were neglected, except inputs of the Broad Channel residential area mentioned as items 15 and 16 above.

SIMULATION

The computer simulation was carried out for three calendar days from May 31, 1972 through June 2, 1972, with $T = 0$ at the 0000 hour of May 31, which is 20 hr before the beginning of the rainstorm. The time step of the simulation was 1 min ($\Delta t = 60$ sec). Several simulations of coliform density distributions were made simultaneously, although only that with the sixteen input sources will be reported here. A coliform disappearance rate of $2.65 \times 10^{-5}$/sec, which corresponds to a 90-percent decay in concentration per day, was used throughout the simulation. It was not considered necessary to make extensive model adjustments for water quality similar to those for the tide. In Vol. III of this study [9] several simulations presented for average conditions in the bay agreed reasonably well with observation.

Also, we did not have available several measurements of the bay’s response to rainstorms to make extensive adjustments in the diffusion coefficients using data from one storm and then from another storm as a final check, in the way that the tide model was adjusted.

Two complete simulations were made; the first was used only as a check upon the adequacy of the simulation input files, because many sources had been added since previous simulations. This simulation was also made for quality control of the graphic output. Results of that simulation indicated inconsistencies in inputs, which were corrected; the second simulation made is presented here. Before the latter simulation was made, all the source input estimates were discussed with City engineers and checked with plant records for these days.

A typical pollutant constituent map resulting from the simulation is shown in Fig. 38. The following information is given in this chart, which represents the situation at 2232 min from the beginning of the simulation:

1. Locations of the sixteen peripheral inputs, indicated by (0).
2. Clock time and in some instances the date.
3. Wind speed and direction.
4. Tidal current velocity vector (1 ft/sec corresponding to a grid) on every other point in each direction.
Fig. 38—A typical pollutant constituent map from the simulation showing synoptic tidal current field, pollutant field, and numerical values of pollutants and water levels at a particular time.

5. Coliform concentrations and tide stage at sixteen stations throughout the bay. The left value is the concentration value (to be multiplied by 1000); the right is the water level.

6. Contours of the coliform concentrations with the values of isolines indicated at the lower right-hand corner of each map. The coliform concentrations shown on the map must be multiplied by 1000. The lowest contour indicated is at 24 MPN/ml.

The two charts of Figs. 39 and 40, time steps 1240 and 1364, show the conditions of the bay 40 min and 2 hr 44 min after the beginning of the storm, respectively. Figure 40 (time step 1364 at 2244 hr May 31) indicates that, after three hours, the impact of the storm is realized only within the creeks of the North Channel, whereas the pollutants have already reached the Broad Channel area and the South Channel near the Rockaway outfall because of their locations and the relatively short system detention time.

The next 22 graphs in Figs. 41 through 62 show the condition of the bay at 2-hr and 4-min intervals covering the period from midnight of June 1 to 8 p.m. of June 2.
Fig. 39—Coliform distribution estimates on May 31, 1972 at 2040, approximately 1 hour after start of rainstorm.
Fig. 40—Coliform distribution estimates on May 31, 1972 at 2244, approximately 3 hours after start of rainstorm
Fig. 41—Coliform distribution estimates on June 1, 1972 at 0048, approximately 5 hours after start of rainstorm
Fig. 42—Coliform distribution estimates on June 1, 1972 at 0252, approximately 7 hours after start of rainstorm.
Fig. 43—Coliform distribution estimates on June 1, 1972 at 0456, approximately 9 hours after start of rainstorm.
Fig. 44—Coliform distribution estimates on June 1, 1972 at 0700, approximately 11 hours after start of rainstorm
Fig. 45—Coliform distribution estimates on June 1, 1972 at 0904, approximately 13 hours after start of rainstorm
Fig. 46—Coliform distribution estimates on June 1, 1972 at 1108, approximately 15 hours after start of rainstorm.
Fig. 47—Coliform distribution estimates on June 1, 1972 at 1312, approximately 17 hours after start of rainstorm
Fig. 48—Coliform distribution estimates on June 1, 1972 at 1516, approximately 19 hours after start of rainstorm
Fig. 50—Coliform distribution estimates on June 1, 1972 at 1924, approximately 23 hours after start of rainstorm
Fig. 51—Coliform distribution estimates on June 1, 1972 at 2128, approximately 25 hours after start of rainstorm
Fig. 54—Coliform distribution estimates on June 2, 1972 at 0340, approximately 32 hours after start of rainstorm.
Fig. 55—Coliform distribution estimates on June 2, 1972 at 0544, approximately 34 hours after start of rainstorm
Fig. 56—Coliform distribution estimates on June 2, 1972 at 0748, approximately 36 hours after start of rainstorm.
Fig. 57—Coliform distribution estimates on June 2, 1972 at 0952, approximately 38 hours after start of rainstorm.
Fig. 58—Coliform distribution estimates on June 2, 1972 at 1156, approximately 40 hours after start of rainstorm.
Fig. 59—Coliform distribution estimates on June 2, 1972 at 1400, approximately 42 hours after start of rainstorm
Fig. 60—Coliform distribution estimates on June 2, 1972 at 1604, approximately 44 hours after start of rainstorm
Fig. 61—Coliform distribution estimates on June 2, 1972 at 1808, approximately 46 hours after start of rainstorm.
The rainstorm started a few hours before HW. Because several hours must elapse before overflow occurs, the initial discharges into the bay are around HW slack; thus the primary mechanism for pollutant transport is the diffusion and advection induced by the local gradient, which in turn is caused by the waste water inputs. (See Fig. 40.) These patterns begin to accelerate and deform as the advection by the tidal currents becomes more important (Fig. 41). As the water level is falling, water from the peripheral creeks is being discharged into the major tidal channels within the bay system (Fig. 42). At this time the horizontal plumes of pollutant from the Rockaway drainage system is quite well established because of the well defined tidal transport pattern in the receiving waters around the outfall. On the other hand, the patterns of pollutant in the areas near Bergen Basin propagate slowly in all directions as a result of small tidal transports in the back of the bay system.

Approximately nine hours after the start of the rainstorm the impact of a combined sewer overflow discharge is felt throughout the entire North Channel area (Fig. 43). At the same time, pollutants discharged from the Rockaway combined sewer overflow are now being carried back toward the east by the flooding tide. However, the total amount of active bacterial load in the bay originating from this source is reduced to approximately half of its original quantity due to the dieaway effect (7 hr of contact time for exponential dieoff with $T_{90} = 24$ hr). Eleven hours after the start of the rainstorm (Fig. 44), pollutants discharged in the North Channel area are carried eastward, while those from Bergen Basin propagate farther in Grassy Bay. The same trend can be seen in Figs. 45 and 46, except that more marsh areas are now being affected when the tidal flats are submerged during high tide.

When the tide starts to ebb again at noon of June 1, the southwesterly wind becomes more pronounced (see the wind speed and direction plot at the lower right portion of map). The wind effect counterbalances some of the ebbing tidal transport, resulting in complicated patterns of pollutant distributions, especially near the Pumpkin Patch area in the center of the bay, where the fetch length is longest for this particular wind direction (Fig. 48). Because of the shallowness of this marsh area, the wind-induced transport almost follows the wind direction, with negligible deflection from the Coriolis effect.

After one cycle of complete tidal excursion, the source of bacteria near the Mill Basin area can no longer be distinguished between the Rockaway or the Mill Basin separate storm systems.

From Fig. 45 and Fig. 52 it can be seen that one day after the storm the water qualities in North Channel are still around 100 MPN/ml. The extent of the area with unacceptable quality (i.e., more than 24 MPN/ml) is now decreasing. In the South Channel most of the areas are now of acceptable quality, the only exception being near the Beach Channel area, where pollutants from the Rockaway discharge did not drop to an acceptable level until about four hours later (Fig. 54).

About 36 hr after the storm, the discharged pollutants in the North Channel and Mill Basin entrance areas are well mixed by the excursive tidal transport and diffusive mechanisms. Therefore, from this time on the reductions of bacterial concentration in this general area are due mainly to the dieoff effects, as can be seen from the sequence of maps in Figs. 55 through 62. It takes approximately two days before the water quality in the bay reaches the standard acceptable for bathing.

The computed temporal variations of coliform concentrations at sixteen selected water-quality stations were also plotted. Five of these histories are presented in Figs. 63 and 64, and the remainder in Appendix C. These coliform concentration histograms and the previously described pollutant distribution maps give essentially the
Fig. 63 — Coliform estimates from observation and simulation at two stations in Jamaica Bay with upper and lower limits of the 95 percent confidence bands, together with the water level history.
Fig. 64—Coliform estimates from observation and simulation at three stations in Jamaica Bay
spatial and temporal distribution of the fate of the waste inputs within the Jamaica Bay system. The observed coliform concentrations at five major water-quality stations are also plotted in Figs. 63 and 64. Maps at the upper left-hand corner of the graphs give the location of each station. In addition, the water level history is described in Fig. 63.

In these figures the 90-percent confidence interval is shown for the observations as well as the computed estimates. The upper limit shown is the upper 95-percent confidence limit, and the lower the lower 95-percent confidence limit. Thus, in a statistical sense, we have 90-percent confidence that the computed and observed estimates are within the interval of the limits shown.

The 95-percent confidence bands for the computed water-quality estimates are derived from the confidence levels of the time histories of urban drainage quantities and qualities hindcasted by the urban drainage models from rainfall patterns. The width of this band is compatible with that of the observed coliform concentrations in the bay because the same number of dilutions were used in the coliform fermentation tests for deriving the coliform prediction in the urban drainage models as described by Liu [2]. The confidence bands for the computed bay water-quality estimates do not, however, include the prediction variances originated from the tidal transport generated from the hindcasted water levels which drive the tidal flows from the model open boundary. The variabilities from this source are small because the confidence bands for the amplitude and phase of frequency response functions estimated from the U.S. Geological Survey gauge at East Rockaway to the reconstructed water level at the model open boundary are small (approximately ±0.002 and ±0.096 degrees angle, respectively).

The estimated concentration histories with confidence bands obtained from the simulation agree well for stations 7 and 8 (Fig. 63), taking into account the variability of the constituent. This is also the case for the second day of the simulation for stations 9 and 10, and the third day for the estimates of station 14 near the entrance of Bergen Basin. However, agreement between computed and measured coliform estimates is not good on the third day of the simulation for stations 9 and 10, and for the second day of the simulation for station 14.

From the time history of coliform estimates of station 14 near Bergen Basin, it is clear that the origin of the discrepancy is underestimated sources in Bergen Basin. Volume IV of this series [4] indicated that a weak counterclockwise circulation exists in the bay. Consequently, the low Bergen Basin estimate appears as an underestimate of the computed coliform values at stations 9 and 10, which are west of station 14, on the third day of the simulation.

If a clockwise circulation had existed, the high coliform levels from this source would not have moved westward; thus we find here an indirect confirmation of the direction of mass transport in the bay.

For the computed histories of water quality at the other eleven stations, presented in Appendix C, no sample data was available for the period of this test.

To analyze the water movements through the bay as well as the movements of coliform, the total volume of water and the total number of bacteria passing through eight cross-sections of the bay were computed. The water volumes are the total from the start of the computation. In Fig. 65, the top graph presents the water transport in a cross-section at the transition of Grassy Bay to Island Channel. The downward trend on the graph indicates that the net movement is toward the west. This is also the case for the total coliform transport. The graph of the diffusive transport indicates initially an eastward transport, about two orders of magnitude smaller than the advective transport.

This diffusion direction is a surprise, since a strong source is located just east
Fig. 65—Water and pollutant transport histories through cross-sections in the northeastern part of Jamaica Bay, together with computed water levels.
of the cross-section and one would expect a net diffusive transport westward. However, from the distribution of coliform at 0700 hr on June 1 (time step 1860) in the vicinity of the cross-section (Figs. 44 and 66), it can be seen that at the confluence of Pumpkin Patch Channel and Island Channel a counterclockwise local circulation is induced around low-water (LW) slack. This circulation caused higher coliform concentration west of the southern part of the cross-section, then east of the southern part of the cross-section; the eastward diffusion apparently prevailed.

Figure 65 presents the computed mass transport and the computed bacterial transport through Broad Beach Channel. The general direction of movement is counterclockwise.

Figures 67 through 69 present transport through other cross-sections, together with local tide information. In all the computed results a counterclockwise circulation is found. However, from the computed results for a section near Canarsie Pier, one can conclude that the net circulation is very small, which also agrees with earlier findings.

The total coliform transport is not always in the same direction as the net water movement. The time of the release, the diffusive transport, and the decay all influence this total transport. Generally the diffusive transport through the cross-sections is about one order of magnitude smaller than the advective transport. The exception is an east-west section in the island channel located just south of the Paerdegat Basin inlet. The combined sewer overflow discharge in this basin causes a very high coliform concentration in Paerdegat Basin, which in turn causes a plume-type discharge into the main channel in the morning of June 1, 1972 (Fig. 45). The cross-section is situated just south of the inlet, causing a very high coliform concentration gradient over the cross-section, which then results in the large diffuse transport of Fig. 67.

In Fig. 69 the transports through a section across Rockaway Inlet are presented. Of interest is the coliform transport through this section. In the computational procedure of the concentration distributions, the concentration at the boundary is determined during ebb from information inside the computational field, such as the local velocity, the concentration, and the concentration gradient. When the current changes direction and flows into the estuary, this procedure can no longer be applied, since the concentration boundary condition is determined outside the computational field. Thus, either new assumptions must be made or measured information should be used. In this case it was assumed that the concentration decreased over a three-hour period starting from LW slack to a value of 10 MPN/ml according to a half-cosine variation. It can be seen in the graph of the transport of the total number of organisms that an eastward, thus inward, transport exists. Our procedures then have the effect of a source existing west of the boundary of a magnitude similar to that of the smaller drainage basin along the north shore of the bay, and several times larger than the emission of the Coney Island WPCF outfall when this facility is chlorinating to 24 MPN/ml. The limited influence of this source is felt only in the computed distributions of the southwest part of the bay during flood.

In all the computed tide records (Figs. 67 through 69), the influence of the pulse type discharge from the overflows can be seen as a perturbation on the record late in the evening of May 31, 1971.
Fig. 66—An enlarged section of constituent map for time 1860 (Fig. 44) near the transition of Grassy Bay to Island Channel, showing the local water movements and the pollutant distribution. See text and Fig. 65 for detailed explanation of the pollutant transport mechanisms.
Fig. 17—Water and coliform transport histories through cross-sections near Canarsie Pier, together with the computed water levels
Fig. 68—Water and coliform transport histories through cross-sections of Pumpkin Patch Channel, together with the computed water levels
Fig. 69—Water and coliform transport histories through cross-sections of Rockaway Inlet and Beach Channel, together with the computed water levels.
V. DISCUSSION

SIMULATION

From the results of the simulation, it is apparent that the inputs into Bergen Basin are underestimated. Several major discharges enter into Bergen Basin, i.e., two separate storm sewer systems (Bergen and Kennedy) and a combined sewer overflow. The characteristics of the combined sewer overflows discharging into Jamaica Bay have been measured extensively, and the generated inputs agree well with the measurement of overflows of similar storms.

The estimates of the separate systems are less accurate, since they are based upon very limited data. For the Bergen (separate) drainage system (Fig. 1) a value for the coliform concentration was estimated at 9000 MPN/ml. This estimate was based upon field samples from the Thurston separate drainage system observed during the summer of 1970. It was assumed that this value would also be representative for this basin.

Following the simulation, a review of a drainage map of this area revealed that there is an overflow from the sanitary system leading to the Jamaica plant into the Bergen Basin storm system, and that this overflow is activated when the Jamaica Bay plant nears its maximum capacity. It is suspected that this transfer from the sanitary system into the separate system is the cause of the high loadings in Bergen Basin, particularly since the graph of the observed flow rate shows a nearly constant level after the rainstorm at approximately the maximum capacity (see left bottom graph in Fig. 34).

There are several other possible causes of the discrepancy between observed and computed values in the bay near Bergen Basin, such as the following:

1. High level of bacterial concentration discharged from the Kennedy Airport storm system, entering the system at the western end of Bergen Basin.
2. Pronounced resuspension of bacteria in Bergen Basin due to the high flow created by three major inflows, i.e., Bergen combined system (2699 acres), Bergen separate system (approximately 7000 acres), and the west portion of Kennedy Airport (approximately 2200 acres).
3. Some degree of aftergrowth of coliform in the Grassy Bay water, which is characterized by its relatively high BOD and low turbulence level.

From the available data it is impossible to delineate exactly the source of the difference between observation and computation, although the most probable source is the overflow from the sanitary system into the Bergen storm system during and after an intense rainstorm.

Since the City plans to treat all discharges of combined sewer overflows, the coliform inputs of the separate systems are of major interest. It is quite possible that these will be the major sources of coliform input after all planned AWPCFs for the combined sewer overflows are constructed. Consequently it is required that the unknown source be determined and that particular attention be given simultaneously to the quality of the discharges of the separate storm sewer system. A monitoring program for tracing the unknown source is presented in Appendix C. The data
generated by that program will also give detailed information for another evaluation by which the reliability of future analyses will be enhanced.*

WATER-QUALITY MANAGEMENT AND CONTROL

The analysis of the simulation results and comparison with field data indicate that the prediction of coliform water quality throughout the bay is possible following a rainstorm. Also, the overflow quantity and quality of drainage basins can be predicted from rainfall. This information is a basic part of the inputs required for the preliminary design of auxiliary water pollution control systems for Jamaica Bay.

During the study reported here it has become apparent that much of the basic data concerning the existing sewer system is not readily available for designing that system. We traced an overflow connection between two systems from a discrepancy between our computed estimates and observation, and it is quite possible that other overflows exist between the different types of systems. Such connections, which will influence the response characteristics of the drainage basin in quality and quantity, should be known before any preliminary design is made. Thus it is recommended that a civil engineering study be made of the existing conditions in the drainage basin areas prior to the studies for the design of future auxiliary water pollution control systems.

Subsequently, a management program should be developed for a system that will incorporate the existing sewer network, new and existing auxiliary treatment plants, and the treatment plants now being upgraded. The system should be analyzed as a whole in order to obtain optimum overall characteristics, such as the dimensions of components in relation to cost and effectiveness. For the latter, an assessment of periods in which certain water-quality standards are achieved at locations of interest is critical.

After the analysis of the system and choice of the most optimal method of treatment for the auxiliary plants, preliminary designs for these facilities can be made.

* The monitoring program and the new evaluation of the model have recently been executed with good results and will be reported in the future.
VI. CONCLUSIONS

From coliform water-quality observations in Jamaica Bay following a rainstorm and the simulation of the coliform water quality, the following conclusions were drawn:

1. Coliform-bacteria distributions in Jamaica Bay, as well as discharges into the bay resulting from rainstorms, can be predicted with confidence from rainfall intensity data and tide data by use of the water-quality simulation model and models of the drainage basins.

2. These models can be used as the basic tool for an optimal design of the future auxiliary treatment system and for the optimal management thereof.

3. An unknown source of pollutants in Bergen Basin led to a discrepancy between observed and simulated coliform estimates in the vicinity of this basin; this should be investigated further by use of the field measurement program described in Appendix D.

From the frequency response analysis of tides and tidal flow, the following conclusions were drawn:

1. The spectra of the computed and observed water levels are similar.

2. The frequency response function can be used as an aid in adjusting the bottom friction in a two-dimensional model.

3. Missing data from a tide gauging station can be reconstructed from a neighboring station if the relationship between those stations has been determined.
Appendix A

RESPONSE FUNCTION DETERMINATION BETWEEN
DENSE AND SPARSE DATA SETS

Periodic field observations of geophysical and biological phenomena in estuaries and coastal seas are often interrupted by equipment constraints and personnel availability. As a result, incomplete data sets must be used for analysis. Unfortunately, present methods of analyzing random data are not applicable, since they are based on continuous records or uninterrupted digitalized data sets. Consequently, the incomplete data set is not efficiently used; generally only a "quick look" analysis is used in which only the subjectivity characteristics of the phenomena are obtained.

Since data collection is the major expense in studies, it seems appropriate to develop methods by which a maximum of information is obtained from a minimum amount of data. As cross-spectra analyses of time series are at present the most powerful tool for establishing the relationship between time-varying phenomena, the cross-spectral analysis method described by Liu [2] should be extended to sparse data sets which are interrupted or have missing data values.

In the method of analysis presented here it is assumed that the digitalized data has the usual properties of being stationary and ergodic.

Central in determining the general dependency of the values of one set of data on the other set is the determination of the cross-covariance function. For a pair of uninterrupted records, the cross-covariance functions of \( x(t) \) and \( y(t) \) for a time lag \( \tau \) is determined by

\[
\gamma_{xy}(\tau) = \lim_{T \to \infty} \frac{1}{T} \int_0^T x(t) y(t + \tau) \, dt
\]  

(A-1)

For estimating the cross-covariance function from digitalized records sampled at even time intervals, a numerical equivalent of Eq. (A-1) is used with an integration over a limited period of time, thus using a finite number of sampled values.

If the sampled sets \( [x], [y], t = 1, 2, 3, \ldots, n \) represents \( x(t) \) and \( y(t) \) sampled at interval \( \tau \) over a total time \( n \tau \), the estimate for the cross-covariance function at the displacement \( k \tau \) is, when the maximum lag used in the analysis is the interval \( m \tau \),

\[
\gamma_{xy}(k) = \frac{1}{n - m} \sum_{t=1}^{t=n-m} x_t y_{t+k}
\]  

(A-2)

and

\[
\gamma_{yx} = \frac{1}{n - m} \sum_{t=1}^{t=n-m} x_{t+k} y_t
\]  

(A-3)

For computation of the cross-spectral density function estimate, the even and odd parts of the cross-covariance function are determined by
\[ \hat{A}(k) = \frac{1}{2}[\gamma_{xx}(k) + \gamma_{yy}(k)] \] (A-4)

\[ \hat{B}(k) = \frac{1}{2}[\gamma_{xy}(k) - \gamma_{yx}(k)] \] (A-5)

from which the co-spectral density function is estimated:

\[ \hat{C}_{xy}(f) = 2\pi \left[ \hat{A}(0) + 2 \sum_{k=1}^{k=m-1} \hat{A}(k) \cos(2\pi ft) + \hat{A}(m) \cos(2\pi fm) \right] \] (A-6)

The quadrature spectral density function is estimated by

\[ \hat{Q}_{xy}(f) = 2\pi \left[ 2 \sum_{k=1}^{k=m-1} \hat{B}(k) \sin(2\pi ft) + \hat{B}(m) \sin(2\pi fm) \right] \] (A-7)

The spectral density function of input and output are determined in a similar manner. If, in Eq. (A-8), the output series is replaced by the input series, we obtain the auto-covariance function of the input,

\[ \gamma_{xx}(k) = \frac{1}{n-m} \sum_{t=1}^{t=n-m} x_t x_{t+k} \] (A-8)

from which the input spectral density function is determined.

\[ \hat{P}_{xx}(k) = 2\pi \left[ \gamma_{xx}(0) + 2 \sum_{k=1}^{k=m-1} \gamma_{xx}(k) \cos(2\pi ft) + \gamma_{xx}(m) \cos(2\pi fm) \right] \] (A-9)

For the analysis of the correlation between a dense data set and a sparse data set it is assumed that n data values \( \{x_i\}, t = 1, 2, 3, \ldots, n \) of a stationary process \( x(t) \) with \( \bar{x} = 0 \) is available and also a sparse data set \( \{y_i\} \) containing only i numbers of the set with n data values, \( \{y_i\}, t = 1, 2, 3, \ldots, n \) is available for analysis. The sparse data points are randomly distributed over the n data values. For purposes of estimating joint functions, the complete set \( \{y_i\} \) will be used with the assumption that the set \( \{y_i\} \) contains all sparse data and all missing data is set at zero.

The estimated cross-covariance function at the displacement \( kt \), where \( k \) is the lag number, is

\[ \gamma_{xy}(k) = \frac{1}{t_{xy}} \sum_{t=1}^{t=n-m} x_t y_{t+k} \quad k = 0,1,2 \ldots m \] (A-10)
where $I_{yx}(k)$ is the number of cross-products $x_{y_{t+k}}$ that are unequal to zero. In estimating the cross-covariance function it is assumed that a zero value does not occur in any of the data.

Similarly, we may define

$$
\gamma_{yx}(k) = \frac{1}{I_{yx}(k)} \sum_{t=1}^{t=n-m} x_{t+k} y_t \quad k = 0, 1, 2, \ldots, m \quad (A-11)
$$

where $I_{yx}(k)$ is the number of cross-products $x_{y_{t+k}}$ that are unequal to zero.

The above approximation of the cross-covariance function of the interrupted records $x(t)$ seems appropriate, since it can be considered that this covariance function is obtained from an approximation for the sample times $t_1$ to $t_2$, $t_3$ to $t_4$, $t_5$, $t_6$, for which the cross-covariance function can be written as follows:

$$
\gamma_{xy}(\tau) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} x(t) y(t + \tau) \, dt + \frac{1}{t_4 - t_3} \int_{t_3}^{t_4} x(t) y(t + \tau) \, dt + \ldots \quad (A-12)
$$

Once the covariance functions are known, the co-spectrum and quadrature spectrum can be computed from Eqs. (A-4), (A-5), (A-6), and (A-7). The estimated spectral density of the input can be determined in the normal manner by use of Eqs. (A-8) and (A-9). We assume that the series $x$ is complete.

The estimated spectrum of the output can be computed in a manner similar to that described for the cross-covariance with missing data.

The response function can be determined from estimates of the cross-spectra and the input spectra according to Eqs. (21) and (22).
Appendix B

EFFECT OF TIDAL GAUGE CLOCK ERRORS ON VERIFICATION RESULTS

Reference 4 describes the adjustment and verification procedure of the hydraulic model. From observation of tide levels and currents over a four-day period, data from the first two days were used to adjust Manning's coefficients and obtain a good agreement between observed and computed water levels. Subsequently, the simulation was extended and again a comparison was made. The analysis in Ref. 4 showed that several of the clock mechanisms of the recorders were advancing or retarding.

The difference between observation and computation showed periodic errors (Fig. B-1). By making corrections for the clock errors, a much better fit is obtained, and the standard deviations between observed and computed levels were reduced. Figures B-2 and B-3 present the observed and computed tide levels after correction of clock errors, and Fig. B-4 presents the differences in levels between adjacent stations. Figure B-3 can be compared with Fig. A-1 of Ref. 4, and Fig. B-4 with Fig. A-2 of that reference.
Fig. B-1—Differences between observed and computed tide level data for five stations in Jamaica Bay, showing errors caused by inaccurate tide stage recorder timing mechanism.
Fig. B-2—Differences between observed and computed tide level data for five stations in Jamaica Bay after correction of errors caused by tide stage recorder timing mechanism.
Fig. B-3—Computed and observed water levels for five stations on 29 and 30 October 1970 after correction for errors in timing
Fig. B-4—Computed and observed differences in water levels between pairs of stations after correction for timing errors.
Appendix C

COLIFORM VARIATIONS IN TIME AT DIFFERENT LOCATIONS IN JAMAICA BAY

In addition to the time histories at five stations in the north channel of the bay presented in Figs. 51 and 52, time histories at eleven other stations were obtained from the simulation. The locations of these stations in the model are shown in Fig. C-1 and on a small insert chart of the bay within the graphs of the time histories (Figs. C-2 through C-5).

These graphs present the estimated value from the computation together with the 90-percent confidence interval.
Fig. C-2—Computed coliform histories at three stations in the southern part of Jamaica Bay
Fig. C-3—Computed coliform histories at three locations in the eastern part of Jamaica Bay.
Fig. C-4—Computed coliform histories at three locations in the northern and central parts of Jamaica Bay
Fig. C-5—Computed coliform histories at two locations in the northeastern part of Jamaica Bay
Appendix D

PROPOSED MONITORING PROGRAM

Field monitoring of transient water-quality parameters after a rainstorm in a complex urban-estuarine system is a difficult task. It is more difficult in the case of Jamaica Bay, which has eighteen major time-variable inputs with a short delay time from rainfall to overflow. Traffic conditions in the event of a rainstorm retard the sampling operation even further. In order to design a realistic monitoring program to record the essential water-quality responses in the bay, contingency plans have to be provided according to the tradeoff between the relative weight and volume of data points. The basic monitoring program is divided into four groups, as discussed below.

PREPARATORY WORK AND INITIATION OF SAMPLING OPERATION

1. At Canarsie Pier a tide gauge should be installed and maintained on a weekly basis. Time correction and gauge height reference should be made.
2. Rain gauges at 26th Ward, St. Alban Naval Hospital, and Spring Creek AWPCF should be maintained on a biweekly basis.
3. Local weather forecasts should be used as an alert for the monitoring operation.
4. The storm to be monitored should be an isolated type with an antecedent dry period of at least two days, which will ensure an insignificant background of coliform in the bay water.
5. The magnitude of rainfall should exceed 0.3 in. so that the impact of combined sewer overflows can be realized at a major portion of the bay. The sampling program should not be carried out before such an amount is measured at the Spring Creek AWPCF. Quantitative Precipitation Forecast (QPF) service may be available through the Metropolitan U.S. Weather Service before the rainstorm.

OBSERVATION OF INPUT PARAMETERS FOR SIMULATION

1. After the sampling operation is initiated, several coliform samples of WPCF effluent and the time histories of flow rate during overflow (and bypass, if any) should be obtained, regardless of their operational mode. This information will be used in the simulation as time-variable inputs to the bay.
2. Samples from the following land-based stations (see Fig. D-1) should be collected 2 hours after the initial amount of 0.3 in. has been reached, and they should be sampled at hourly intervals for a duration of 5 to 6 hours.
   - Paerdegat Basin entrance
   - Fresh Creek entrance
   - Bergen Basin separate system (double barrel at the eastern end of Bergen Basin)
- Kennedy storm water (single barrel at the eastern end of Bergen Basin)
- Thurston separate system (double barrel at the end of Thurston Basin).

This sampling is of lower priority than the first four items.

**BAY WATER-QUALITY OBSERVATIONS**

This operation should be initiated the morning following the day of the rainstorm.

1. It is estimated that two boats will be needed for bay water sampling. North Channel stations (7, 8, 9, 10, and 14) can be covered with one boat, while the four South Channel stations can be covered by the other. If only one boat is available, it should also cover the sampling at stations 7a and 2; stations 3 and 4 should be dropped from sampling under this circumstance.

2. Samples collected at stations 7a and 2 should provide a basis for adjusting the pollutant emission rate from the Rockaway combined system overflow and its initial impact on the water quality at the proposed beach sites on the east side of Barren Island. Samples collected on the second and third day after the storm from station 7a would help to assess the total bacterial emission from the North Channel stations, due to the sharp gradients existing near those stations. Sharp gradients make the observed and computed bacterial concentrations very sensitive to the spatial sampling variability. Samples from Bay stations should be taken at two-hour intervals during daylight hours only (≈14 hr per day) for a total duration of 3 days from the beginning of bay water sampling.

**LABORATORY TESTS**

1. Presumptive test of total coliform should be made on all samples.
2. Confirmed tests on some selected samples are desirable only if manpower permits.
3. Selected tests on fecal coliform are also desirable only if manpower permits. If intermittent fecal-coliform tests on selected samples are performed, they should be selected on the following basis:

- First few samples from land-based stations, which enable us to estimate the emission rate from both combined and separate storm systems. Highest levels of both fecal-coliform and fecal streptococcus were found in samples from the entrance of Bergen Basin.
- Last three days' samples from station 14, which enables us to determine the disappearance rate, if it exists. This can be achieved by examining the time histories of concentrations with known diffusive characteristics from a given location.
- The last few samples from station 7a, which enable us to estimate the total emission from North Channel inputs.

DATA COLLECTION AND MANAGEMENT

1. Copies of wind data and rainfall records from the Kennedy Airport weather station should be obtained.
2. Rain gauge charts from three New York City-operated stations should be collected. Time corrections, if any, should be marked on charts before they are taken off the gauge.
3. Tide gauge records should be collected with time corrections and reference points marked.
4. Laboratory test results should be reported in the standard New York City Bacteriological Examination format.

CONCLUDING REMARKS

The purpose of the second monitoring program is twofold, i.e., to determine the extent and origin of unknown inputs into Jamaica Bay from the Bergen Basin area and to obtain inputs for a second check on the adequacy of simulation results from the two-dimensional water quality simulation model. The data obtained from the field and the simulations will considerably enhance the confidence with which predictions of water-quality conditions in Jamaica Bay can be made, thus permitting a more effective design of the system of treatment facilities.
REFERENCES


