US Army Corps of Engineers
Hydrologic Engineering Center

Annual Extreme Lake Elevations by Total Probability Theorem

May 1990

TP-134

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
May 1990	Technical Paper	

4. TITLE AND SUBTITLE
Annual Extreme Lake Elevations by Total Probability Theorem

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Harold E. Kubik, P.E.

5d. PROJECT NUMBER

5e. TASK NUMBER

5F. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center (HEC)
609 Second Street
Davis, CA 95616-4687

8. PERFORMING ORGANIZATION REPORT NUMBER
TP-134

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/ MONITOR'S ACRONYM(S)

11. SPONSOR/ MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
Presented at the Great Lakes Water Level Forecasting and Statistics Symposium, Windsor, Ontario, Canada, 17-18 May 1990.

14. ABSTRACT
Annual extreme water levels on the Great Lakes, whether maximums or minimums, have a high serial dependence. Therefore, application of traditional frequency analysis techniques must be interpreted in a different manner and more sophisticated statistical techniques must be applied to account for this dependence. Decomposition of the annual extremes into two parts, one containing the highly dependent part and the other containing the random part, is one method of dealing with the dependence in the lake elevations. Appropriate statistical analyses can be applied to the separate parts and then the individual results combined to obtain the final frequency relation. This study develops mean monthly lake elevation duration curves to represent the dependent part and wind setup frequency curves for the random part. These parts are then combined by application of the total probability theorem.

15. SUBJECT TERMS
lake elevation, storm surge, total probability theorem, statistics, coincident frequency, Lake Erie

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	18	19b. TELEPHONE NUMBER

Annual Extreme Lake Elevations by Total Probability Theorem

May 1990

US Army Corps of Engineers
Institute for Water Resources
Hydrologic Engineering Center
609 Second Street
Davis, CA 95616

(530) 756-1104
(530) 756-8250 FAX
www.hec.usace.army.mil

TP-134

Papers in this series have resulted from technical activities of the Hydrologic Engineering Center. Versions of some of these have been published in technical journals or in conference proceedings. The purpose of this series is to make the information available for use in the Center's training program and for distribution with the Corps of Engineers.

ANNUAL EXTREME LAKE ELEVATIONS BY TOTAL PROBABILITY THEOREM

*Harold E. Kubik, P.E.**

ABSTRACT: Annual extreme water levels on the Great Lakes, whether maximums or minimums, have a high serial dependence. Therefore, application of traditional frequency analysis techniques must be interpreted in a different manner and more sophisticated statistical techniques must be applied to account for this dependence.

The terms "Percent Chance Exceedance" and "Return Period" are applied to the expectation values of annual extreme events that are random in nature and have an equal likelihood of occurring in any given year. Annual extreme lake elevations on the Great Lakes are not random from one year to the next; therefore, the usual terms to define the expectation should not be used to describe the events. An acceptable term is "Percent of Years Exceeded." This is comparable to the label "Percent to Time Exceeded" that is applied to flow- or elevation-duration curves.

Decomposition of the annual extremes into two parts, one containing the highly dependent part and the other containing the random part, is one method of dealing with the dependence in the lake elevations. Appropriate statistical analyses can be applied to the separate parts and then the individual results combined to obtain the final frequency relation. This study develops mean monthly lake elevation duration curves to represent the dependent part and wind setup frequency curves for the random part. These parts are then combined by application of the total probability theorem.

Seasonality of the occurrence of both parts was found to be very important. Therefore, the complete analysis was done for the six-month fall-winter period and the six-month spring-summer period. The two curves were combined by the union of probabilities.

This technique does not gain any information over a smooth curve drawn through the observed events when applied to long-record gauges like Cleveland and Buffalo harbor. This technique is most useful in application to short-record stations. The long record of monthly lake elevations for a particular lake provides the information for the highly dependent part. The wind setup information for a short-record gauge may be correlated with a nearby long-record gauge to be made more indicative of a longer record.

Application of this method to the Buffalo harbor and Cleveland gauges resulted in computed "1% of Years Exceeded" elevations of 579.79 feet (176.72 meters) and 574.72 feet (175.17 meters) (IGLD 1955), respectively.

Introduction

The Great Lakes are an important natural resource that have attracted a variety of human activities — waterborne commerce, water supply, hydroelectric power, recreation, and habitation — to mention some of the more important ones. The wise management of the lakes and the land adjacent to these bodies of water requires some anticipation of the likely lake levels. The establishment of non-building zones, for instance, relies on an estimate of the likely maximum water levels. Planners and designers

*Research Hydraulic Engineer, U.S. Army Corps of Engineers, Hydrologic Engineering Center, Davis, CA
Reprinted from: Proceedings of the Great Lakes Water Level Forecasting and Statistics Symposium, May 17 & 18, 1990, Windsor, Ontario, Canada. Great Lakes Commission, Ann Arbor, Michigan.

involved in the location of boat harbors and depth of navigation channels need information on the expected minimum water levels. The computation of these likely levels is complicated by the long-term fluctuations of the Great Lakes' water levels.

The normal procedure of establishing zones that are subject to flooding, especially in riverine conditions, is to compute a frequency curve based on the available flood data. One of the requirements for a frequency analysis is that the events are random, independent events. The Great Lakes' water level data *do not* meet this requirement. The annual extreme values are highly correlated from year-to-year because of the strong dependence on the mean level during the year. Therefore, normal frequency analysis procedures can not be applied to these data. It is possible to use statistical analysis techniques to analyze the extremes by separating each event into two components: one the long-time scale, highly dependent fluctuation represented by mean lake elevations; and the second the short-time scale, very independent fluctuations generally caused by wind stress on the lake. These components, after individual analysis, can be recombined to provide an indication of the percent of annual instantaneous maximum events that will exceed a given elevation. Application of these techniques to the annual minimums would provide the percent of annual events that do not exceed (nonexceedance) a given elevation.

Data Available for Analysis

Very long records, by usual hydrologic standards in the U.S., of mean monthly water levels on Lake Erie have been observed at the Cleveland and Buffalo harbor gauges. The Cleveland record is continuous since January 1860 (129 years through 1988). And, although some mean monthly values were recorded for the 1860-1869 period, the continuous record at Buffalo harbor began in March 1887 (nearly 102 years through 1988). A continuous record of annual instantaneous extremes are available for the period 1900-1988 at Buffalo harbor and for period 1904-1988 at Cleveland. Figure 1 is a plot of mean annual lake elevations at Cleveland. One could conclude from this plot that the 129 years of information

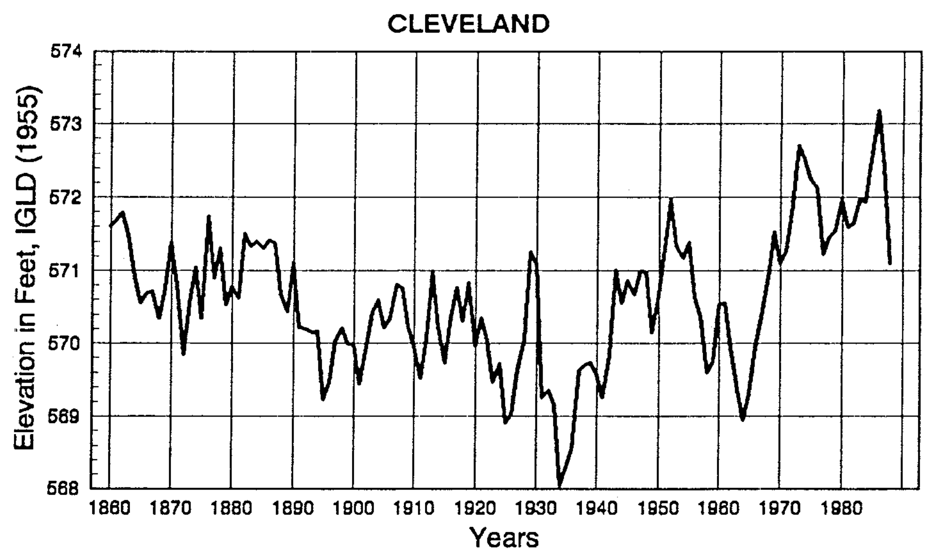

Figure 1. Mean annual elevations on Lake Erie, Cleveland gauge.

is really a very short period. The water levels in the 1860's began fairly high and gradually moved downward until the dramatic decrease in the early 1930s to a low in 1934. After this lowest annual level, the levels generally increased to the high experienced in 1986. Fitting the mean annual elevations with a smooth curve makes it appear that only one-half of a cycle has been observed. The high persistence has effectively reduced our knowledge of how often to expect extreme high or low water levels.

Annual Persistence

Computation of the serial correlation coefficient for the annual extremes, a measure of how well one year is related to the next year, provides a quantitative evaluation of persistence. The lag 1 correlations for the annual maximum events are 0.752 and 0.406 for Cleveland and Buffalo harbor, respectively. The strength of this persistence becomes more clear when it is noted that lags 1 through 4 (this year is related to 4 years previous) are found to be significant.

Comparison of a time series plot of the annual instantaneous extremes, Figures 2 and 3, with the mean annual values illustrate that the extremes have the same pattern as the mean annual values.

As the general lake levels are a large component of the annual extreme, then removal of this component could result in values that *do* meet the frequency requirement of being random and independent. This separation was accomplished by noting the month of the extreme, and subtracting the mean monthly water level at the gauge from the instantaneous extreme. This provided a change in elevation value that is termed "wind setup." (Note, wind setup is negative for the annual instantaneous minimums.) Serial correlation computations indicate that the wind setup values are random events; therefore, frequency analysis techniques can be applied to these data. This provides one component of the annual extreme values.

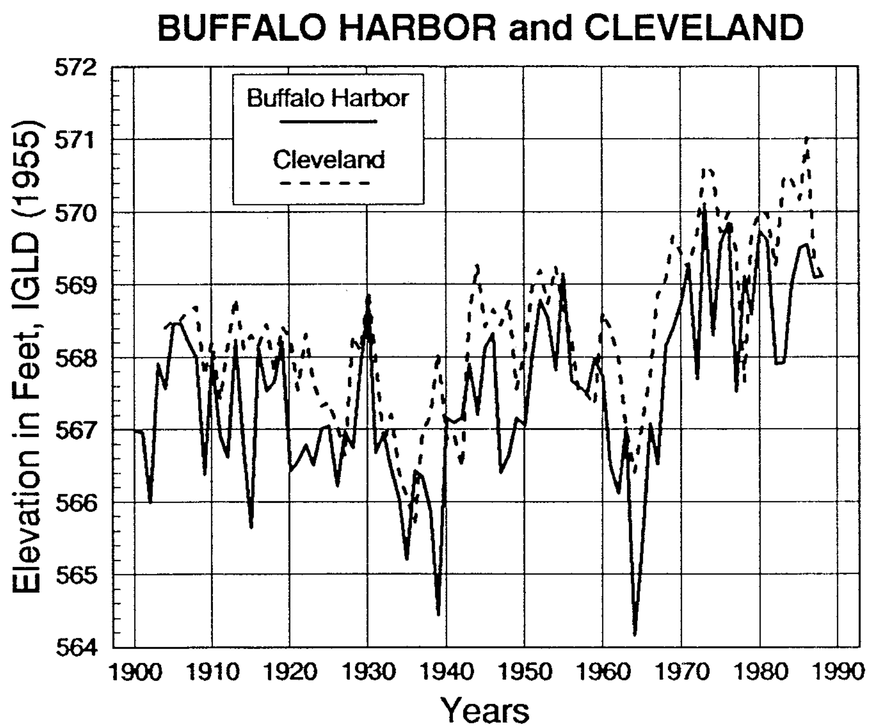

Figure 2. Annual instantaneous maximums at Buffalo harbor and Cleveland.

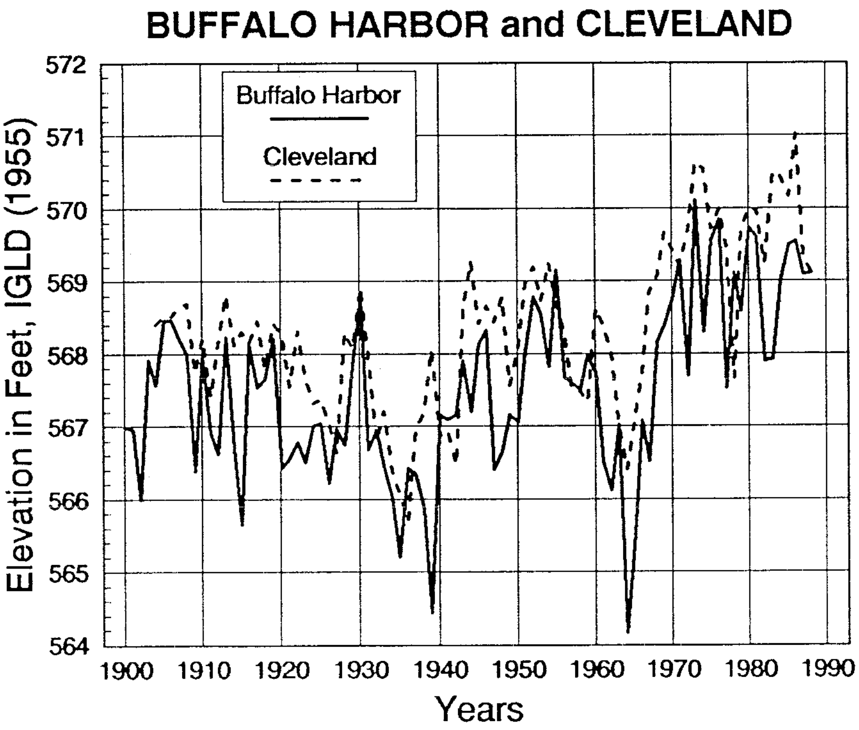

Figure 3. Annual instantaneous minimums at Buffalo harbor and Cleveland.

A second component is the long-term lake fluctuations. This component is represented by a mean monthly elevation duration curve. These values are highly correlated, so the frequency label would be "Percent of Time Exceeded" to imply that they are not independent events.

Seasonality of Extremes

It became apparent as this study progressed that seasonality was important in the analysis of the extreme events. The Buffalo harbor and Cleveland maximum levels occur at entirely different times of the year. The Buffalo harbor maximums occur in the fall-winter months, indicating a response to the winter storms because the monthly lake levels are usually lower during the winter months. At Cleveland, the maximums occur in the spring-summer months indicating that the seasonal high mean lake levels are the larger determining factor. This is illustrated in Figure 4 for the maximum and minimum values at Buffalo harbor and in Figure 5 for Cleveland. For this study, the data were divided into two 6-month seasons. The fall-winter season included the months of October, November, December, January, February, and March. The spring-summer season included the months of April, May, June, July, August and September.

The minimum levels are more influenced by the mean monthly lake levels, although the effect of wind related minimums can be noted at the Buffalo harbor gauge for March and April (February has the lowest average monthly elevation at both gauges).

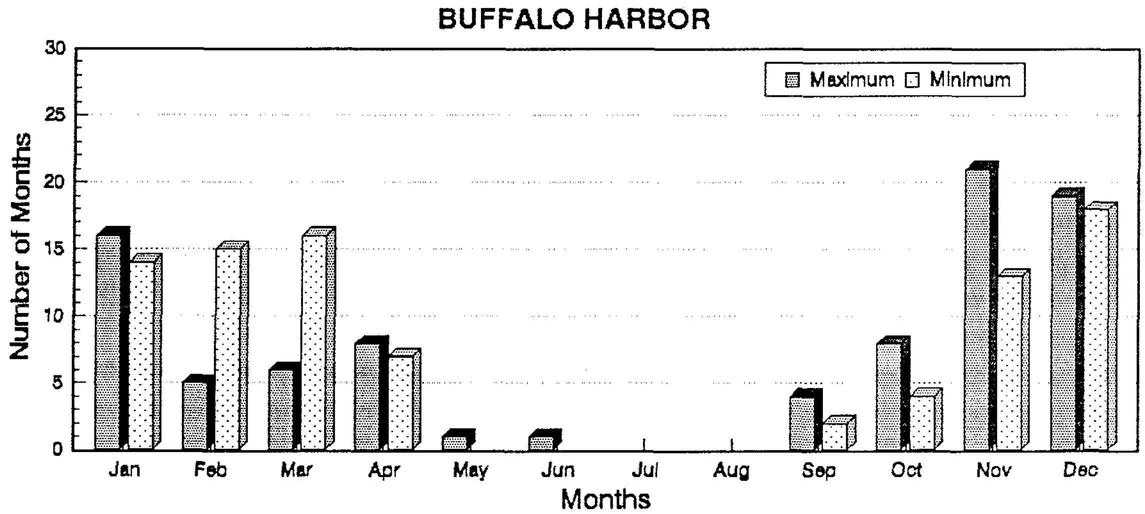

Figure 4. Months of annual maximums and minimums, Buffalo harbor.

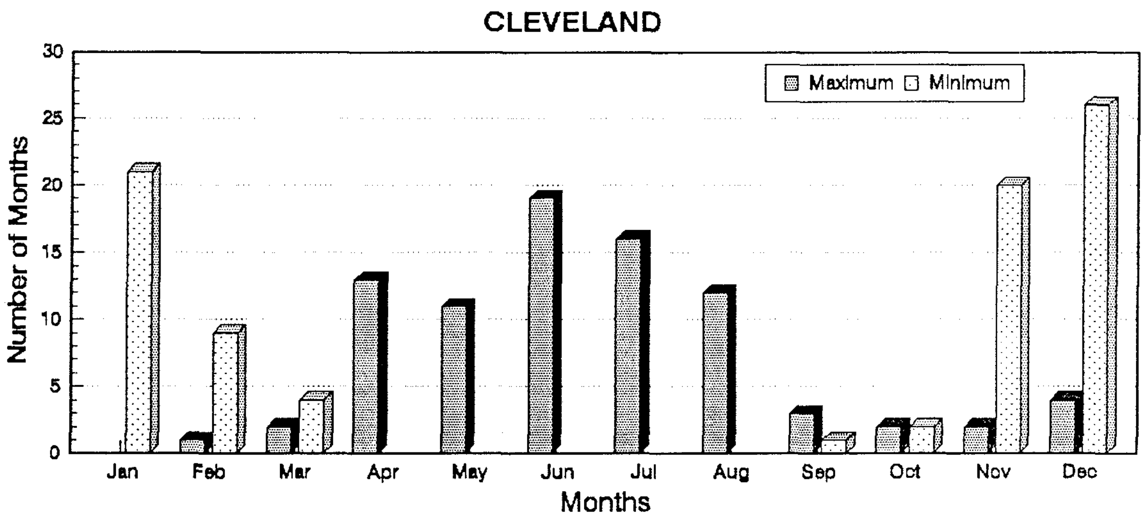

Figure 5. Months of annual maximums and minimums, Cleveland.

Total Probability Method

Now that the annual extremes have been decomposed into two components for each of the seasons, some method must be applied to put the data back together again. This can be done by applying the total probability theorem. The total probability theorem, as presented in most statistics texts (Benjamin and Cornell 1970) is:

$$P[A] = \sum_{i=1}^{n} P[A \mid B_i]\, P[B_i]$$

where:

$P[A \mid B]$ is the conditional probability of the event A given that event B has occurred, and

B is a set of mutually exclusive, collectively exhaustive events of size n.

The conditional probability relations are derived by selecting a given lake elevation and then adding this value to the wind setup frequency curve. This gives a single conditional frequency curve that has a certain probability of occurring. Many of these conditional frequency curves can be computed to completely define the range of water level occurrences. Figure 6 shows seven such conditional frequency curves. Each curve is labeled with the mean monthly lake elevation used to derive the curve and the percent of time that this elevation is exceeded. The horizontal axis (Percent of Years Exceeded) is the P[A|B] portion of the total probability equation. The P[B] portion of the equation is the amount of probability (percent of time) represented by each curve. This can, simplistically, be the probability computed by adding one-half of the differences between the two adjacent curves. For example, the probability associated with the curve based on a monthly elevation of 571.06 (exceeded 50% of the time) would be [(70%-50%)/2 + (50%-30%)/2]/100 = 0.20 units of probability. Doing this for all the curves will yield a set of values that add up to 1.0. In other words, all the possible mean monthly elevations have been considered by discrete increments of probability.

The total probability equation is applied at each desired elevation to compute an expectation of that elevation being exceeded. To derive a frequency relation, several elevations would be selected covering the expected range of values. Figure 7 illustrates in a graphical way what the equation is doing. An elevation of 574.0 was selected, then the Percent of Years Exceeded for each curve is noted and plotted on Figure 7 against the Percent of Time (converted to probability by dividing by 100). After all of the intercepts have been plotted, a smooth curve is drawn through the points. (Note that not all of the curves used to develop Figure 7 are shown on Figure 6.) For an elevation of 574.0, the expected Percent of Years Exceeded of 4.37% is the probability weighted average, or the area under this curve.

This computational procedure is often called coincident frequency analysis in Corps of Engineers publications. As these computations are laborious, a computer program has been written (HEC 1989) that accepts as input the mean monthly elevation-duration relation and the wind setup frequency relation. The program then generates the requisite conditional curves and evaluates the total probability theorem for several elevations to provide an elevation expectation relation.

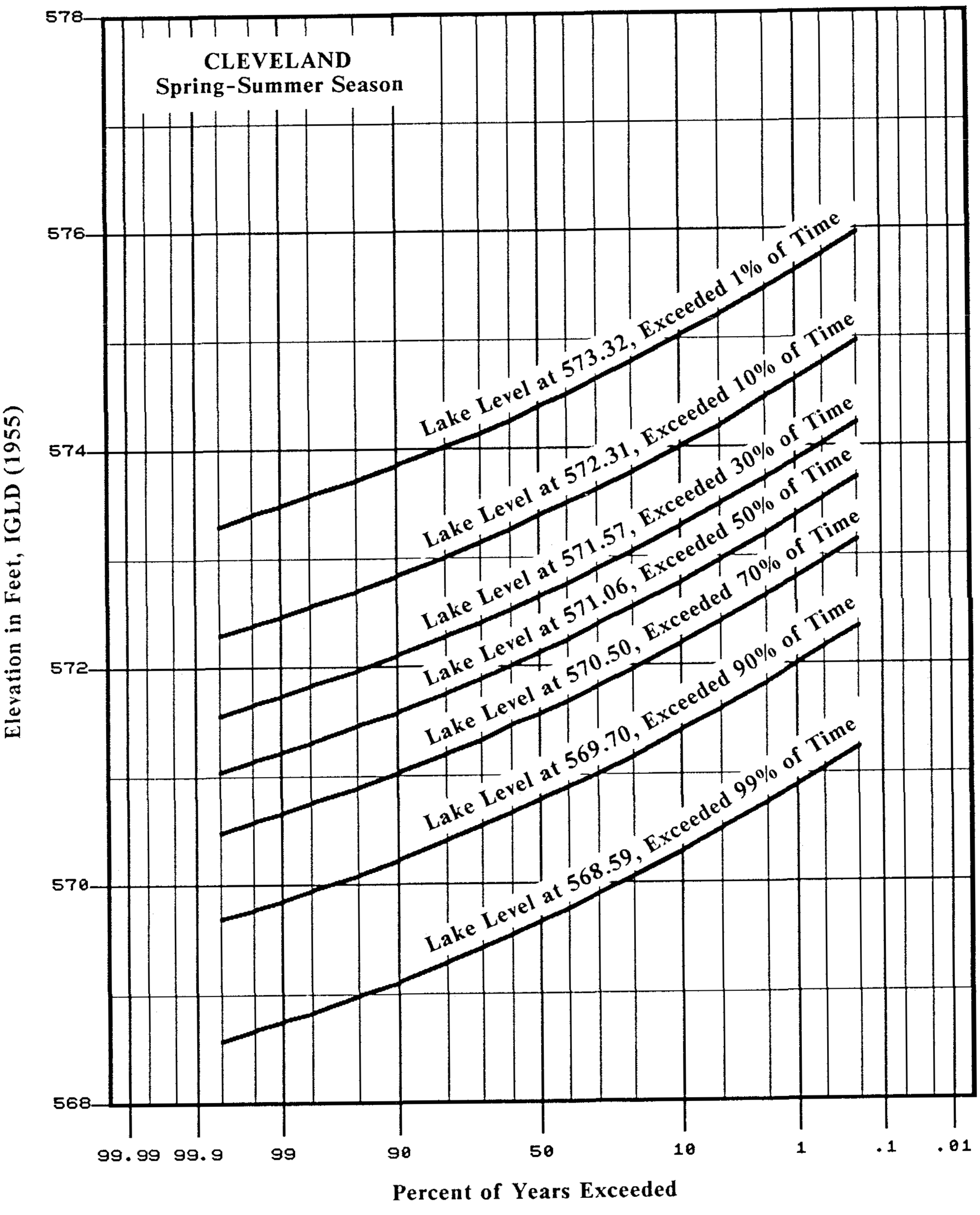

Figure 6. Conditional frequency curves, Cleveland, spring-summer season.

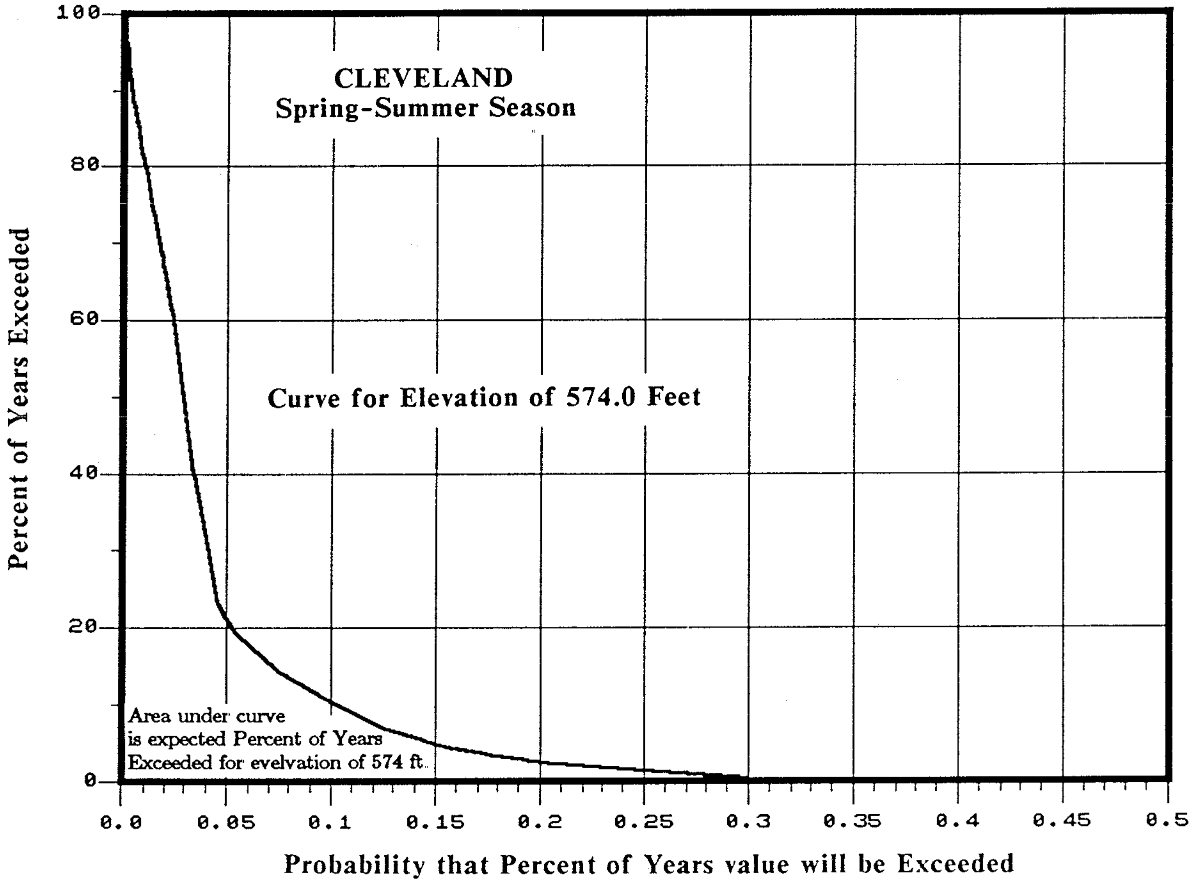

Figure 7. Graphical representation of applying the total probability equation.

Results

The final results were found by combining the computed "frequency curves" for each of the seasons. This is done by the union of probabilities. This equation is:

$$P_c = 100[1 - (1 - P_1/100)(1 - P_2/100)]$$

where: P_c = the combined frequency value in percent for the selected elevation,

P_1 = the frequency value in percent for season 1 for selected elevation, and

P_2 = the frequency value in percent for season 2 for selected elevation.

Lake elevation expectation curves were computed for Buffalo harbor and Cleveland by the procedure described herein. The monthly duration curves were based on the period 1860-1988 while the wind setup curves were based generally on the 1900-1988 period. Therefore, these curves should be fairly representative of the 1860-1988 period. The observed instantaneous annual maximums have been assigned plotting positions and plotted along with the derived curves on Figures 8 and 9. The "1% of

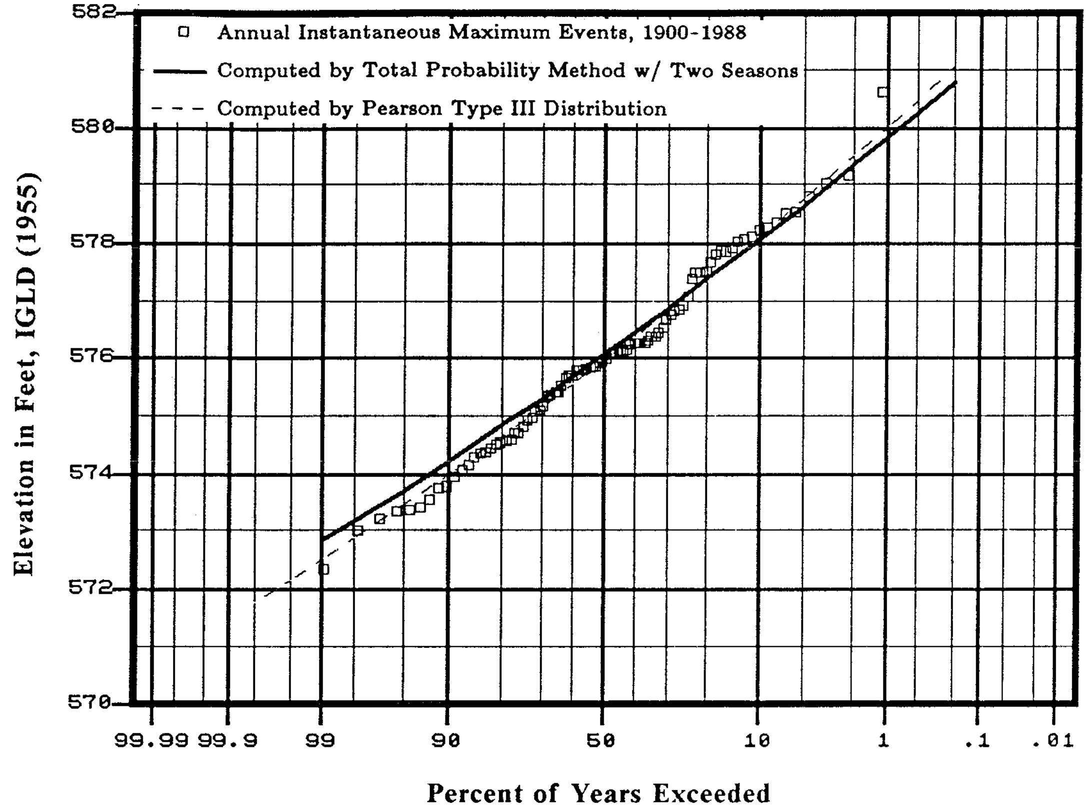

Figure 8. Frequency of annual maximums, Buffalo harbor.

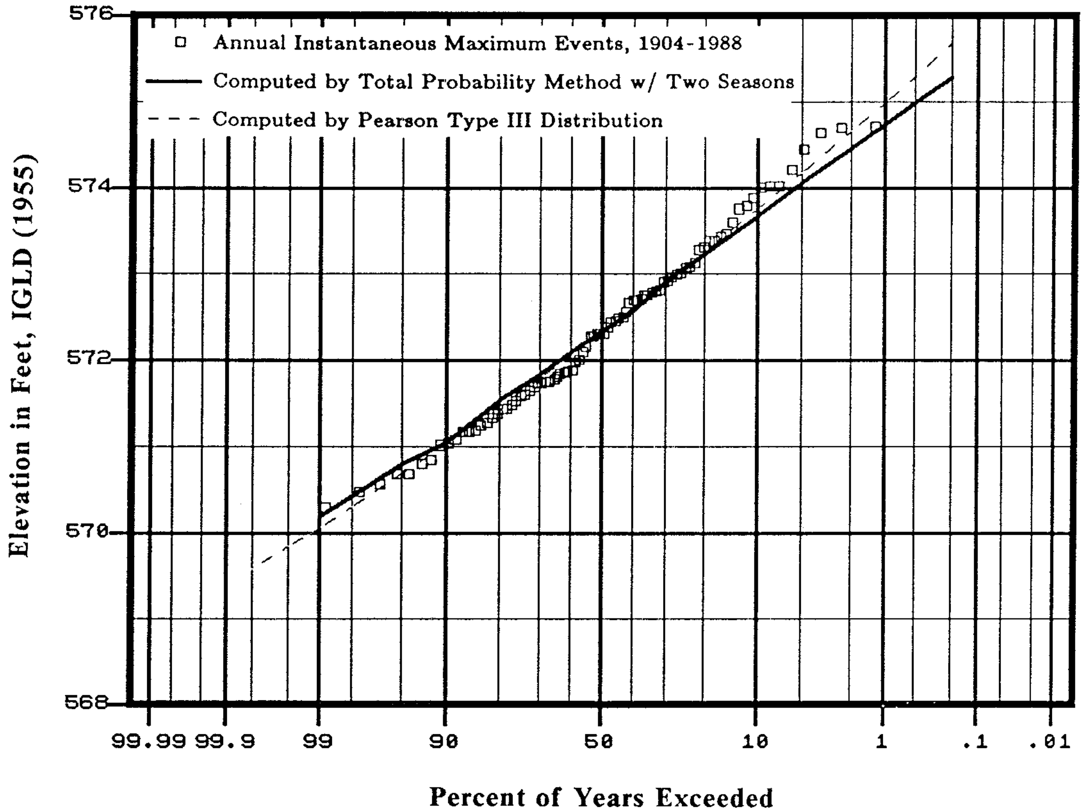

Figure 9. Frequency of annual maximums, Cleveland.

Years Exceeded" elevations computed by this procedure were 579.79 (176.72 meters) and 574.72 feet (175.17 meters) (IGLD 1955) for Buffalo harbor and Cleveland, respectively.

The utility of this procedure is in the application to gauges that have fairly short records. Mean monthly elevation duration relations based on a fairly long period are available for each of the Great Lakes. The wind setup frequency relation for an individual station may be used, or the relation could be adjusted by the "two-station comparison" procedures (Interagency Committee 1982) recommended for flood flow frequency computations. Application of these procedures to a station with a fairly short record should provide elevation expectation curves that are representative of a much longer period than the period of recorded maximum or minimum instantaneous lake elevations.

Acknowledgments

This work was supported in part by the U.S. Army Corps of Engineers, Detroit District and the research program of the Hydrologic Engineering Center. This work was done under the supervision of Arlen Feldman, Chief Research Division and Darryl Davis, Director. I wish to thank the Great Lakes Environmental Laboratory for inviting me to participate in this exercise in water level analysis.

References

Benjamin, J. R. and C. A. Cornell. 1970. *Probability, Statistics, and Decision for Civil Engineers*. New York, NY: McGraw-Hill Book Co.

Interagency Advisory Committee on Water Data. 1982. *Guidelines for Determining Flood Flow Frequency*, Bulletin 17B. Reston, VA: Hydrology Subcommittee.

U.S. Army Corps of Engineers. 1989. *CFA, Coincident Frequency Analysis*. Computer Program. Davis, CA: Hydrologic Engineering Center.

Technical Paper Series

TP-1 Use of Interrelated Records to Simulate Streamflow
TP-2 Optimization Techniques for Hydrologic Engineering
TP-3 Methods of Determination of Safe Yield and Compensation Water from Storage Reservoirs
TP-4 Functional Evaluation of a Water Resources System
TP-5 Streamflow Synthesis for Ungaged Rivers
TP-6 Simulation of Daily Streamflow
TP-7 Pilot Study for Storage Requirements for Low Flow Augmentation
TP-8 Worth of Streamflow Data for Project Design - A Pilot Study
TP-9 Economic Evaluation of Reservoir System Accomplishments
TP-10 Hydrologic Simulation in Water-Yield Analysis
TP-11 Survey of Programs for Water Surface Profiles
TP-12 Hypothetical Flood Computation for a Stream System
TP-13 Maximum Utilization of Scarce Data in Hydrologic Design
TP-14 Techniques for Evaluating Long-Tem Reservoir Yields
TP-15 Hydrostatistics - Principles of Application
TP-16 A Hydrologic Water Resource System Modeling Techniques
TP-17 Hydrologic Engineering Techniques for Regional Water Resources Planning
TP-18 Estimating Monthly Streamflows Within a Region
TP-19 Suspended Sediment Discharge in Streams
TP-20 Computer Determination of Flow Through Bridges
TP-21 An Approach to Reservoir Temperature Analysis
TP-22 A Finite Difference Methods of Analyzing Liquid Flow in Variably Saturated Porous Media
TP-23 Uses of Simulation in River Basin Planning
TP-24 Hydroelectric Power Analysis in Reservoir Systems
TP-25 Status of Water Resource System Analysis
TP-26 System Relationships for Panama Canal Water Supply
TP-27 System Analysis of the Panama Canal Water Supply
TP-28 Digital Simulation of an Existing Water Resources System
TP-29 Computer Application in Continuing Education
TP-30 Drought Severity and Water Supply Dependability
TP-31 Development of System Operation Rules for an Existing System by Simulation
TP-32 Alternative Approaches to Water Resources System Simulation
TP-33 System Simulation of Integrated Use of Hydroelectric and Thermal Power Generation
TP-34 Optimizing flood Control Allocation for a Multipurpose Reservoir
TP-35 Computer Models for Rainfall-Runoff and River Hydraulic Analysis
TP-36 Evaluation of Drought Effects at Lake Atitlan
TP-37 Downstream Effects of the Levee Overtopping at Wilkes-Barre, PA, During Tropical Storm Agnes
TP-38 Water Quality Evaluation of Aquatic Systems
TP-39 A Method for Analyzing Effects of Dam Failures in Design Studies
TP-40 Storm Drainage and Urban Region Flood Control Planning
TP-41 HEC-5C, A Simulation Model for System Formulation and Evaluation
TP-42 Optimal Sizing of Urban Flood Control Systems
TP-43 Hydrologic and Economic Simulation of Flood Control Aspects of Water Resources Systems
TP-44 Sizing Flood Control Reservoir Systems by System Analysis
TP-45 Techniques for Real-Time Operation of Flood Control Reservoirs in the Merrimack River Basin
TP-46 Spatial Data Analysis of Nonstructural Measures
TP-47 Comprehensive Flood Plain Studies Using Spatial Data Management Techniques
TP-48 Direct Runoff Hydrograph Parameters Versus Urbanization
TP-49 Experience of HEC in Disseminating Information on Hydrological Models
TP-50 Effects of Dam Removal: An Approach to Sedimentation
TP-51 Design of Flood Control Improvements by Systems Analysis: A Case Study
TP-52 Potential Use of Digital Computer Ground Water Models
TP-53 Development of Generalized Free Surface Flow Models Using Finite Element Techniques
TP-54 Adjustment of Peak Discharge Rates for Urbanization
TP-55 The Development and Servicing of Spatial Data Management Techniques in the Corps of Engineers
TP-56 Experiences of the Hydrologic Engineering Center in Maintaining Widely Used Hydrologic and Water Resource Computer Models
TP-57 Flood Damage Assessments Using Spatial Data Management Techniques
TP-58 A Model for Evaluating Runoff-Quality in Metropolitan Master Planning
TP-59 Testing of Several Runoff Models on an Urban Watershed
TP-60 Operational Simulation of a Reservoir System with Pumped Storage
TP-61 Technical Factors in Small Hydropower Planning
TP-62 Flood Hydrograph and Peak Flow Frequency Analysis
TP-63 HEC Contribution to Reservoir System Operation
TP-64 Determining Peak-Discharge Frequencies in an Urbanizing Watershed: A Case Study
TP-65 Feasibility Analysis in Small Hydropower Planning
TP-66 Reservoir Storage Determination by Computer Simulation of Flood Control and Conservation Systems
TP-67 Hydrologic Land Use Classification Using LANDSAT
TP-68 Interactive Nonstructural Flood-Control Planning
TP-69 Critical Water Surface by Minimum Specific Energy Using the Parabolic Method

TP-142 Systems Analysis Applications at the Hydrologic Engineering Center
TP-143 Runoff Prediction Uncertainty for Ungauged Agricultural Watersheds
TP-144 Review of GIS Applications in Hydrologic Modeling
TP-145 Application of Rainfall-Runoff Simulation for Flood Forecasting
TP-146 Application of the HEC Prescriptive Reservoir Model in the Columbia River Systems
TP-147 HEC River Analysis System (HEC-RAS)
TP-148 HEC-6: Reservoir Sediment Control Applications
TP-149 The Hydrologic Modeling System (HEC-HMS): Design and Development Issues
TP-150 The HEC Hydrologic Modeling System
TP-151 Bridge Hydraulic Analysis with HEC-RAS
TP-152 Use of Land Surface Erosion Techniques with Stream Channel Sediment Models
TP-153 Risk-Based Analysis for Corps Flood Project Studies - A Status Report
TP-154 Modeling Water-Resource Systems for Water Quality Management
TP-155 Runoff simulation Using Radar Rainfall Data
TP-156 Status of HEC Next Generation Software Development
TP-157 Unsteady Flow Model for Forecasting Missouri and Mississippi Rivers
TP-158 Corps Water Management System (CWMS)
TP-159 Some History and Hydrology of the Panama Canal
TP-160 Application of Risk-Based Analysis to Planning Reservoir and Levee Flood Damage Reduction Systems
TP-161 Corps Water Management System - Capabilities and Implementation Status